

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 June 2003 (12.06.2003)

PCT

(10) International Publication Number
WO 03/047526 A2

(51) International Patent Classification⁷: **A61K**

(21) International Application Number: PCT/US02/38437

(22) International Filing Date:
26 November 2002 (26.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|------------|-------------------------------|----|
| 60/334,343 | 30 November 2001 (30.11.2001) | US |
| 60/340,278 | 7 December 2001 (07.12.2001) | US |
| 60/345,069 | 4 January 2002 (04.01.2002) | US |
| 60/351,352 | 25 January 2002 (25.01.2002) | US |
| 60/357,168 | 14 February 2002 (14.02.2002) | US |
| 60/369,128 | 29 March 2002 (29.03.2002) | US |
| 60/370,802 | 5 April 2002 (05.04.2002) | US |

(71) Applicant (for all designated States except US): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **BECHA, Shanya, D.** [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). **BHATIA, Umesh** [US/US]; 5212 Union Avenue, San Jose, CA 95124 (US). **BLAKE, Julie, J.** [US/US]; 3818 Pacheco Street, San Francisco, CA 94116 (US). **BOROWSKY, Mark, L.** [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). **BURRILL, John, D.** [US/US]; 2218 Brewster Avenue, Redwood City, CA 94062 (US). **DELEGEANE, Angelo, M.** [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). **ELLIOTT, Vicki, S.** [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). **GANDHI, Ameena, R.** [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). **GIETZEN, Kimberly, J.** [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). **GORVAD, Ann, E.** [US/US]; 369 Marie Common, Livermore, CA 94550 (US). **GRIFFIN, Jennifer, A.** [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). **HO, Anne** [FR/US]; 1279 Poplar Avenue, #114, Sunnyvale, CA 94086 (US).

JIN, Pei [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US). **KABLE, Amy, E.** [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). **LAL, Preeti, G.** [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **LEE, Ernestine, A.** [US/US]; 20523 Crow Creek Road, Castro Valley, CA 94552 (US). **LEE, Sally** [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). **LEE, Soo, Yeun** [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). **MARQUIS, Joseph, P.** [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). **LEHR-MASON, Patricia, M.** [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **RICHARDSON, Thomas, W.** [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). **SPRAGUE, William, W.** [US/US]; 611 13th Street # C, Sacramento, CA 95814 (US). **SWARNAKAR, Anita** [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). **TANG, Tom, Y.** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **TRAN, Bao** [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). **TRAN, Uyen, K.** [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). **CHAWLA, Narinder, K.** [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). **WARREN, Bridget, A.** [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). **XU, Yuming** [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **ZHENG, Wenjin** [CN/US]; 9 Sutter Creek Lane, Mountain View, CA 94043 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,

[Continued on next page]

(54) Title: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

(57) Abstract: Various embodiments of the invention provide human cell adhesion and extracellular matrix proteins (CADECM) and polynucleotides which identify and encode CADECM. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CADECM.



WO 03/047526 A2



ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *without international search report and to be republished upon receipt of that report*

CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

TECHNICAL FIELD

The invention relates to novel nucleic acids, cell adhesion and extracellular matrix proteins encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of immune system disorders, neurological disorders, developmental disorders, connective tissue disorders, and cell proliferative disorders, including cancer. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and cell adhesion and extracellular matrix proteins.

BACKGROUND OF THE INVENTION

Cell Adhesion Proteins

The surface of a cell is rich in transmembrane proteoglycans, glycoproteins, glycolipids, and receptors. These macromolecules mediate adhesion with other cells and with components of the ECM. The interaction of the cell with its surroundings profoundly influences cell shape, strength, flexibility, motility, and adhesion. These dynamic properties are intimately associated with signal transduction pathways controlling cell proliferation and differentiation, tissue construction, and embryonic development. Families of cell adhesion molecules include the cadherins, integrins, lectins, neural cell adhesion proteins, and some members of the proline-rich proteins.

Cadherins comprise a family of calcium-dependent glycoproteins that function in mediating cell-cell adhesion in virtually all solid tissues of multicellular organisms. These proteins share multiple repeats of a cadherin-specific motif, and the repeats form the folding units of the cadherin extracellular domain. Cadherin molecules cooperate to form focal contacts, or adhesion plaques, between adjacent epithelial cells. The cadherin family includes the classical cadherins and protocadherins. Classical cadherins include the E-cadherin, N-cadherin, and P-cadherin subfamilies. E-cadherin is present on many types of epithelial cells and is especially important for embryonic development. N-cadherin is present on nerve, muscle, and lens cells and is also critical for embryonic development. P-cadherin is present on cells of the placenta and epidermis. Recent studies report that protocadherins are involved in a variety of cell-cell interactions (Suzuki, S.T. (1996) J. Cell Sci. 109:2609-2611). The intracellular anchorage of cadherins is regulated by their dynamic association with catenins, a family of cytoplasmic signal transduction proteins associated with the actin cytoskeleton. The anchorage of cadherins to the actin cytoskeleton appears to be regulated by protein tyrosine phosphorylation, and the cadherins are the target of phosphorylation-induced junctional disassembly (Aberle, H. et al. (1996) J. Cell. Biochem. 61:514-523).

Integrins are ubiquitous transmembrane adhesion molecules that link the ECM to the internal

cytoskeleton. Integrins are composed of two noncovalently associated transmembrane glycoprotein subunits called α and β . At least 8 different β subunits ($\beta 1$ - $\beta 8$) and at least 12 different α subunits have been identified ($\alpha 1$ - $\alpha 8$, αL , αM , αX , and αIIb). Individual α subunits are capable of associating with different β subunits, suggesting a possible mechanism for specifying integrin function and ligand binding affinity. Members of the β subunit family are generally of 90-110 kilodaltons (kD) in molecular weight and share about 40-48% amino acid sequence homology. About 56 cysteines distributed among four repeating units are also conserved. Some variation in these conserved features is observed among some of the more divergent β subunit family members. Members of the α subunit family are generally 150-200 kilodaltons in molecular weight and are not as well conserved as the β subunit family. All contain seven repeating domains of 24-45 amino acids spaced about 20-35 amino acids apart. The N-termini each contain 3-4 divalent cation binding sites. (For review, see Pigott, R. and C. Power (1994) The Adhesion Molecule Facts Book, Academic Press, San Diego, CA, pp. 9-12.)

Integrins function as receptors that specifically recognize and bind to ECM proteins such as fibronectin, fibrinogen, laminin, thrombospondin, vitronectin, von Willebrand factor, and collagen. Some integrins recognize a specific motif, the RGD sequence, at the C-termini of the ECM proteins they bind. Integrins also bind to immunoglobulin superfamily proteins such as ICAM-1, -2, and -3 and VCAM-1.

Most integrins have been shown to activate focal adhesion kinase (FAK), a protein tyrosine kinase that is linked to Ras signaling pathways that modify the cytoskeleton and stimulate the mitogen-activated protein kinase (MAPK) cascade (Hanks, S.K. and T.R. Polte (1997) *BioEssays* 19:137-145). Integrins can also influence growth factor signaling through direct interaction with growth factor receptor tyrosine kinases (RTKs) (Miyamoto, S. et al. (1996) *J. Cell Biol.* 135:1633-1642). Integrins have also been shown to play a vital role in "anoikis," a term describing programmed cell death caused by loss of cell anchorage (Frisch, S.M. and E. Ruoslahti (1997) *Curr. Opin. Cell Biol.* 9:701-706).

A number of diseases have been attributed to integrin defects. (See Pigott and Power, *supra*). For example, leukocyte adhesion deficiency (LAD) is an inherited disorder characterized by the impaired migration of neutrophils to sites of extravascular inflammation. LAD is caused by abnormal splicing of and a missense mutation in the RNA encoding the $\beta 2$ subunit. Additionally, defects in platelet integrin are correlated with Glanzmann's thrombasthenia, a bleeding disorder characterized by insufficient platelet aggregation.

Regulators of Integrin Function

A number of integrin-binding proteins which regulate integrin-cytoskeleton and integrin-ECM interactions have been identified. For example, cytohesin-1, a $\beta 2$ subunit binding protein,

enhances integrin-ECM interactions when overexpressed (Kolanus, W. et al. (1996) Cell 86:233-242). Cytohesin-1 resembles nucleotide exchange factors and lipid binding proteins, suggesting that cytohesin-1 may relay intracellular signals to membrane-bound integrins. Additionally, calreticulin, an α subunit binding protein, has been shown to promote integrin-mediated cell adhesion, possibly by modulating intracellular Ca^{2+} levels (Coppolino, M.G. et al. (1997) Nature 386:843-847).

Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (reviewed in Drickamer, K. and Taylor, M. E. (1993) Annu. Rev. Cell Biol. 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L. A. (1991) J. Cell. Biochem. 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

G-Protein Signaling

Guanine nucleotide binding proteins (G-proteins) are critical mediators of signal transduction between a particular class of extracellular receptors, the G-protein coupled receptors (GPCRs), and intracellular second messengers such as cAMP and Ca^{2+} . G-proteins are linked to the cytosolic side of a GPCR such that activation of the GPCR by ligand binding stimulates binding of the G-protein to GTP, inducing an "active" state in the G-protein. In the active state, the G-protein acts as a signal to trigger other events in the cell such as the increase of cAMP levels or the release of Ca^{2+} into the cytosol from the ER, which, in turn, regulate phosphorylation and activation of other intracellular proteins. Recycling of the G-protein to the inactive state involves hydrolysis of the bound GTP to GDP by a GTPase activity in the G-protein. (See Alberts, B. et al. (1994) Molecular Biology of the Cell Garland Publishing, Inc. New York, NY, pp.734-759.) The superfamily of G-proteins consists of several families which may be grouped as translational factors, heterotrimeric G-proteins involved in transmembrane signaling processes, and low molecular weight (LMW) G-proteins including the proto-oncogene Ras proteins and products of rab, rap, rho, rac, smg21, smg25, YPT, SEC4, and ARF genes, and tubulins (Kaziro, Y. et al. (1991) Annu. Rev. Biochem. 60:349-400). In all cases, the GTPase activity is regulated through interactions with other proteins.

Heterotrimeric G-proteins are composed of 3 subunits, α , β , and γ , which in their inactive conformation associate as a trimer at the inner face of the plasma membrane. $\text{G}\alpha$ binds GDP or GTP and contains the GTPase activity. The $\beta\gamma$ complex enhances binding of $\text{G}\alpha$ to a receptor. $\text{G}\gamma$ is necessary for the folding and activity of $\text{G}\beta$ (Neer, E.J. et al. (1994) Nature 371:297-300). Multiple homologs of each subunit have been identified in mammalian tissues, and different combinations of subunits have specific functions and tissue specificities (Spiegel, A.M. (1997) J. Inher. Metab. Dis. 20:113-121).

The alpha subunits of heterotrimeric G-proteins can be divided into four distinct classes. The

α -s class is sensitive to ADP-ribosylation by pertussis toxin which uncouples the receptor:G-protein interaction. This uncoupling blocks signal transduction to receptors that decrease cAMP levels which normally regulate ion channels and activate phospholipases. The inhibitory α -I class is also susceptible to modification by pertussis toxin which prevents α -I from lowering cAMP levels. Two
5 novel classes of α subunits refractory to pertussis toxin modification are α -q, which activates phospholipase C, and α -12, which has sequence homology with the *Drosophila* gene concertina and may contribute to the regulation of embryonic development (Simon, M.I. (1991) *Science* 252:802-808).

The mammalian G β and G γ subunits, each about 340 amino acids long, share more than 80%
10 homology. The G β subunit (also called transducin) contains seven repeating units, each about 43 amino acids long. The activity of both subunits may be regulated by other proteins such as calmodulin and phosducin or the neural protein GAP 43 (Clapham, D. and E. Neer (1993) *Nature* 365:403-406). The β and γ subunits are tightly associated. The β subunit sequences are highly conserved between species, implying that they perform a fundamentally important role in the
15 organization and function of G-protein linked systems (Van der Voorn, L. (1992) *FEBS Lett.* 307:131-134). They contain seven tandem repeats of the WD-repeat sequence motif, a motif found in many proteins with regulatory functions. WD-repeat proteins contain from four to eight copies of a loosely conserved repeat of approximately 40 amino acids which participates in protein-protein interactions. Mutations and variant expression of β transducin proteins are linked with various
20 disorders. Mutations in LIS1, a subunit of the human platelet activating factor acetylhydrolase, cause Miller-Dieker lissencephaly. RACK1 binds activated protein kinase C, and RbAp48 binds retinoblastoma protein. CstF is required for polyadenylation of mammalian pre-mRNA in vitro and associates with subunits of cleavage-stimulating factor. Defects in the regulation of β -catenin contribute to the neoplastic transformation of human cells. The WD40 repeats of the human F-box
25 protein bTrCP mediate binding to β -catenin, thus regulating the targeted degradation of β -catenin by ubiquitin ligase (Neer, supra; Hart, M. et al. (1999) *Curr. Biol.* 9:207-210). The γ subunit primary structures are more variable than those of the β subunits. They are often post-translationally modified by isoprenylation and carboxyl-methylation of a cysteine residue four amino acids from the C-terminus; this appears to be necessary for the interaction of the $\beta\gamma$ subunit with the membrane and
30 with other G-proteins. The $\beta\gamma$ subunit has been shown to modulate the activity of isoforms of adenylyl cyclase, phospholipase C, and some ion channels. It is involved in receptor phosphorylation via specific kinases, and has been implicated in the p21ras-dependent activation of the MAP kinase cascade and the recognition of specific receptors by G-proteins (Clapham and Neer, supra).

G-proteins interact with a variety of effectors including adenylyl cyclase (Clapham and Neer,
35 supra). The signaling pathway mediated by cAMP is mitogenic in hormone-dependent endocrine

tissues such as adrenal cortex, thyroid, ovary, pituitary, and testes. Cancers in these tissues have been related to a mutationally activated form of a $G\alpha_s$ known as the gsp (Gs protein) oncogene (Dhanasekaran, N. et al. (1998) *Oncogene* 17:1383-1394). Another effector is phosducin, a retinal phosphoprotein, which forms a specific complex with retinal $G\beta$ and $G\gamma$ ($G\beta\gamma$) and modulates the ability of $G\beta\gamma$ to interact with retinal $G\alpha$ (Clapham and Neer, supra).

Irregularities in the G-protein signaling cascade may result in abnormal activation of leukocytes and lymphocytes, leading to the tissue damage and destruction seen in many inflammatory and autoimmune diseases such as rheumatoid arthritis, biliary cirrhosis, hemolytic anemia, lupus erythematosus, and thyroiditis. Abnormal cell proliferation, including cyclic AMP stimulation of brain, thyroid, adrenal, and gonadal tissue proliferation is regulated by G proteins. Mutations in $G\alpha$ subunits have been found in growth-hormone-secreting pituitary somatotroph tumors, hyperfunctioning thyroid adenomas, and ovarian and adrenal neoplasms (Meij, J.T.A. (1996) *Mol. Cell Biochem.* 157:31-38; Aussel, C. et al. (1988) *J. Immunol.* 140:215-220).

LMW G-proteins are GTPases which regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. They consist of single polypeptides which, like the alpha subunit of the heterotrimeric G-proteins, are able to bind to and hydrolyze GTP, thus cycling between an inactive and an active state. LMW G-proteins respond to extracellular signals from receptors and activating proteins by transducing mitogenic signals involved in various cell functions. The binding and hydrolysis of GTP regulates the response of LMW G-proteins and acts as an energy source during this process (Bokoch, G.M. and C.J. Der (1993) *FASEB J.* 7:750-759).

At least sixty members of the LMW G-protein superfamily have been identified and are currently grouped into the ras, rho, arf, sar1, ran, and rab subfamilies. Activated ras genes were initially found in human cancers, and subsequent studies confirmed that ras function is critical in determining whether cells continue to grow or become differentiated. Ras1 and Ras2 proteins stimulate adenylate cyclase (Kaziro, supra), affecting a broad array of cellular processes. Stimulation of cell surface receptors activates Ras which, in turn, activates cytoplasmic kinases. These kinases translocate to the nucleus and activate key transcription factors that control gene expression and protein synthesis (Barbacid, M. (1987) *Annu. Rev. Biochem.* 56:779-827, Treisman, R. (1994) *Curr. Opin. Genet. Dev.* 4:96-98). Other members of the LMW G-protein superfamily have roles in signal transduction that vary with the function of the activated genes and the locations of the G-proteins that initiate the activity. Rho G-proteins control signal transduction pathways that link growth factor receptors to actin polymerization, which is necessary for normal cellular growth and division. The rab, arf, and sar1 families of proteins control the translocation of vesicles to and from membranes for protein processing, localization, and secretion. Vesicle- and target- specific identifiers (v-SNAREs and t-SNAREs) bind to each other and dock the vesicle to the acceptor membrane. The budding

process is regulated by the closely related ADP ribosylation factors (ARFs) and SAR proteins, while rab proteins allow assembly of SNARE complexes and may play a role in removal of defective complexes (Rothman, J. and F. Wieland (1996) Science 272:227-234). Ran G-proteins are located in the nucleus of cells and have a key role in nuclear protein import, the control of DNA synthesis, and cell-cycle progression (Hall, A. (1990) Science 249:635-640; Barbacid, M. supra; Ktistakis, N. (1998) BioEssays 20:495-504; and Sasaki, T. and Y. Takai (1998) Biochem. Biophys. Res. Commun. 245:641-645).

Rab proteins have a highly variable amino terminus containing membrane-specific signal information and a prenylated carboxy terminus which determines the target membrane to which the Rab proteins anchor. More than 30 Rab proteins have been identified in a variety of species, and each has a characteristic intracellular location and distinct transport function. In particular, Rab1 and Rab2 are important in ER-to-Golgi transport; Rab3 transports secretory vesicles to the extracellular membrane; Rab5 is localized to endosomes and regulates the fusion of early endosomes into late endosomes; Rab6 is specific to the Golgi apparatus and regulates intra-Golgi transport events; Rab7 and Rab9 stimulate the fusion of late endosomes and Golgi vesicles with lysosomes, respectively; and Rab10 mediates vesicle fusion from the medial Golgi to the trans Golgi. Mutant forms of Rab proteins are able to block protein transport along a given pathway or alter the sizes of entire organelles. Therefore, Rabs play key regulatory roles in membrane trafficking (Schimmöller, I.S. and S.R. Pfeffer (1998) J. Biol. Chem. 243:22161-22164).

The function of Rab proteins in vesicular transport requires the cooperation of many other proteins. Specifically, the membrane-targeting process is assisted by a series of escort proteins (Khosravi-Far, R. et al. (1991) Proc. Natl. Acad. Sci. USA 88:6264-6268). In the medial Golgi, it has been shown that GTP-bound Rab proteins initiate the binding of VAMP-like proteins of the transport vesicle to syntaxin-like proteins on the acceptor membrane, which subsequently triggers a cascade of protein-binding and membrane-fusion events. After transport, GTPase-activating proteins (GAPs) in the target membrane are responsible for converting the GTP-bound Rab proteins to their GDP-bound state. And finally, guanine-nucleotide dissociation inhibitor (GDI) recruits the GDP-bound proteins to their membrane of origin.

The cycling of LMW G-proteins between the GTP-bound active form and the GDP-bound inactive form is regulated by a variety of proteins. Guanosine nucleotide exchange factors (GEFs) increase the rate of nucleotide dissociation by several orders of magnitude, thus facilitating release of GDP and loading with GTP. The best characterized is the mammalian homolog of the Drosophila Son-of-Sevenless protein. Certain Ras-family proteins are also regulated by guanine nucleotide dissociation inhibitors (GDIs), which inhibit GDP dissociation. The intrinsic rate of GTP hydrolysis of the LMW G-proteins is typically very slow, but it can be stimulated by several orders of

magnitude by GAPs (Geyer, M. and A. Wittinghofer (1997) *Curr. Opin. Struct. Biol.* 7:786-792). Both GEF and GAP activity may be controlled in response to extracellular stimuli and modulated by accessory proteins such as RalBP1 and POB1. Mutant Ras-family proteins, which bind but cannot hydrolyze GTP, are permanently activated, and cause cell proliferation or cancer, as do GEFs that inappropriately activate LMW G-proteins, such as the human oncogene NET1, a Rho-GEF (Drivas, G.T. et al. (1990) *Mol. Cell Biol.* 10:1793-1798; Alberts, A.S. and R. Treisman (1998) *EMBO J.* 14:4075-4085).

A member of the ARF family of G-proteins is centaurin beta 1A, a regulator of membrane traffic and the actin cytoskeleton. The centaurin β family of GTPase-activating proteins (GAPs) and Arf guanine nucleotide exchange factors contain pleckstrin homology (PH) domains which are activated by phosphoinositides. PH domains bind phosphoinositides, implicating PH domains in signaling processes. Phosphoinositides have a role in converting Arf-GTP to Arf-GDP via the centaurin β family and a role in Arf activation (Kam, J.L. et al. (2000) *J. Biol. Chem.* 275:9653-9663). The rho GAP family is also implicated in the regulation of actin polymerization at the plasma membrane and in several cellular processes. The gene ARHGAP6 encodes GTPase-activating protein 6 isoform 4. Mutations in ARHGAP6, seen as a deletion of a 500 kb critical region in Xp22.3, causes the syndrome microphthalmia with linear skin defects (MLS). MLS is an X-linked dominant, male-lethal syndrome (Prakash, S.K. et al. (2000) *Hum. Mol. Genet.* 9:477-488).

A member of the Rho family of G-proteins is CDC42, a regulator of cytoskeletal rearrangements required for cell division. CDC42 is inactivated by a specific GAP (CDC42GAP) that strongly stimulates the GTPase activity of CDC42 while having a much lesser effect on other Rho family members. CDC42GAP also contains an SH3-binding domain that interacts with the SH3 domains of cell signaling proteins such as p85 alpha and c-Src, suggesting that CDC42GAP may serve as a link between CDC42 and other cell signaling pathways (Barfod, E.T. et al. (1993) *J. Biol. Chem.* 268:26059-26062).

The Dbl proteins are a family of GEFs for the Rho and Ras G-proteins (Whitehead, I.P. et al. (1997) *Biochim. Biophys. Acta* 1332:F1-F23). All Dbl family members contain a Dbl homology (DH) domain of approximately 180 amino acids, as well as a pleckstrin homology (PH) domain located immediately C-terminal to the DH domain. Most Dbl proteins have oncogenic activity, as demonstrated by the ability to transform various cell lines, consistent with roles as regulators of Rho-mediated oncogenic signaling pathways. The kalirin proteins are neuron-specific members of the Dbl family, which are located to distinct subcellular regions of cultured neurons (Johnson, R.C. (2000) *J. Cell Biol.* 275:19324-19333).

Other regulators of G-protein signaling (RGS) also exist that act primarily by negatively regulating the G-protein pathway by an unknown mechanism (Druey, K.M. et al. (1996) *Nature*

379:742-746). Some 15 members of the RGS family have been identified. RGS family members are related structurally through similarities in an approximately 120 amino acid region termed the RGS domain and functionally by their ability to inhibit the interleukin (cytokine) induction of MAP kinase in cultured mammalian 293T cells (Druey et al., *supra*).

5 The Immuno-associated nucleotide (IAN) family of proteins has GTP-binding activity as indicated by the conserved ATP/GTP-binding site P-loop motif. The IAN family includes IAN-1, IAN-4, IAP38, and IAG-1. IAN-1 is expressed in the immune system, specifically in T cells and thymocytes. Its expression is induced during thymic events (Poirier, G.M.C. et al. (1999) J. Immunol. 163:4960-4969). IAP38 is expressed in B cells and macrophages and its expression is
10 induced in splenocytes by pathogens. IAG-1, which is a plant molecule, is induced upon bacterial infection (Krucken, J. et al. (1997) Biochem. Biophys. Res. Commun. 230:167-170). IAN-4 is a mitochondrial membrane protein which is preferentially expressed in hematopoietic precursor 32D cells transfected with wild-type versus mutant forms of the bcr/abl oncogene. The bcr/abl oncogene is known to be associated with chronic myelogenous leukemia, a clonal myelo-proliferative disorder,
15 which is due to the translocation between the bcr gene on chromosome 22 and the abl gene on chromosome 9. Bcr is the breakpoint cluster region gene and abl is the cellular homolog of the transforming gene of the Abelson murine leukemia virus. Therefore, the IAN family of proteins appears to play a role in cell survival in immune responses and cellular transformation (Daheron, L. et al. (2001) Nucleic Acids Res. 29:1308-1316).

20 Formin-related genes (FRL) comprise a large family of morphoregulatory genes and have been shown to play important roles in morphogenesis, embryogenesis, cell polarity, cell migration, and cytokinesis through their interaction with Rho family small GTPases. Formin was first identified in mouse limb deformity (*ld*) mutants where the distal bones and digits of all limbs are fused and reduced in size. FRL contains formin homology domains FH1, FH2, and FH3. The FH1 domain has
25 been shown to bind the Src homology 3 (SH3) domain, WWP/WW domains, and profilin. The FH2 domain is conserved and was shown to be essential for formin function as disruption at the FH2 domain results in the characteristic *ld* phenotype. The FH3 domain is located at the N-terminus of FRL, and is required for associating with Rac, a Rho family GTPase (Yayoshi-Yamamoto, S. et al. (2000) Mol. Cell. Biol. 20:6872-6881).

30 Lectins comprise a ubiquitous family of extracellular glycoproteins which bind cell surface carbohydrates specifically and reversibly, resulting in the agglutination of cells (reviewed in Drickamer, K. and M.E. Taylor (1993) Annu. Rev. Cell Biol. 9:237-264). This function is particularly important for activation of the immune response. Lectins mediate the agglutination and mitogenic stimulation of lymphocytes at sites of inflammation (Lasky, L.A. (1991) J. Cell. Biochem.
35 45:139-146; Paietta, E. et al. (1989) J. Immunol. 143:2850-2857).

Lectins are further classified into subfamilies based on carbohydrate-binding specificity and other criteria. The galectin subfamily, in particular, includes lectins that bind β -galactoside carbohydrate moieties in a thiol-dependent manner (reviewed in Hadari, Y.R. et al. (1998) J. Biol. Chem. 270:3447-3453). Galectins are widely expressed and developmentally regulated. Galectins contain a characteristic carbohydrate recognition domain (CRD). The CRD comprises about 140 amino acids and contains several stretches of about 1 - 10 amino acids which are highly conserved among all galectins. A particular 6-amino acid motif within the CRD contains conserved tryptophan and arginine residues which are critical for carbohydrate binding. The CRD of some galectins also contains cysteine residues which may be important for disulfide bond formation. Secondary structure predictions indicate that the CRD forms several β -sheets.

Galectins play a number of roles in diseases and conditions associated with cell-cell and cell-matrix interactions. For example, certain galectins associate with sites of inflammation and bind to cell surface immunoglobulin E molecules. In addition, galectins may play an important role in cancer metastasis. Galectin overexpression is correlated with the metastatic potential of cancers in humans and mice. Moreover, anti-galectin antibodies inhibit processes associated with cell transformation, such as cell aggregation and anchorage-independent growth (see, for example, Su, Z.-Z. et al. (1996) Proc. Natl. Acad. Sci. USA 93:7252-7257).

Selectins, or LEC-CAMs, comprise a specialized lectin subfamily involved primarily in inflammation and leukocyte adhesion (Reviewed in Lasky, *supra*). Selectins mediate the recruitment of leukocytes from the circulation to sites of acute inflammation and are expressed on the surface of vascular endothelial cells in response to cytokine signaling. Selectins bind to specific ligands on the leukocyte cell membrane and enable the leukocyte to adhere to and migrate along the endothelial surface. Binding of selectin to its ligand leads to polarized rearrangement of the actin cytoskeleton and stimulates signal transduction within the leukocyte (Brenner, B. et al. (1997) Biochem. Biophys. Res. Commun. 231:802-807; Hadari, K. I. et al. (1997) J. Biol. Chem. 272:28750-28756). Members of the selectin family possess three characteristic motifs: a lectin or carbohydrate recognition domain; an epidermal growth factor-like domain; and a variable number of short consensus repeats (scr or "sushi" repeats) which are also present in complement regulatory proteins.

Neural cell adhesion proteins (NCAPs) play roles in the establishment of neural networks during development and regeneration of the nervous system (Uyemura, K. et al. (1996) Essays Biochem. 31:37-48; Brummendorf, T., and F.G. Rathjen (1996) Curr. Opin. Neurobiol. 6:584-593). NCAP participates in neuronal cell migration, cell adhesion, neurite outgrowth, axonal fasciculation, pathfinding, synaptic target-recognition, synaptic formation, myelination and regeneration. NCAPs are expressed on the surfaces of neurons associated with learning and memory. Mutations in genes encoding NCAPS are linked with neurological diseases, including hereditary neuropathy,

Charcot-Marie-Tooth disease, Dejerine-Sottas disease, X-linked hydrocephalus, MASA syndrome (mental retardation, aphasia, shuffling gait and adducted thumbs), and spastic paraplegia type I. In some cases, expression of NCAP is not restricted to the nervous system. L1, for example, is expressed in melanoma cells and hematopoietic tumor cells where it is implicated in cell spreading and migration, and may play a role in tumor progression (Montgomery, A.M. et al. (1996) J. Cell Biol. 132:475-485).

NCAPs have at least one immunoglobulin constant or variable domain (Uyemura et al., *supra*). They are generally linked to the plasma membrane through a transmembrane domain and/or a glycosyl-phosphatidylinositol (GPI) anchor. The GPI linkage can be cleaved by GPI phospholipase C. Most NCAPs consist of an extracellular region made up of one or more immunoglobulin domains, a membrane spanning domain, and an intracellular region. Many NCAPs contain post-translational modifications including covalently attached oligosaccharide, glucuronic acid, and sulfate. NCAPs fall into three subgroups: simple-type, complex-type, and mixed-type. Simple-type NCAPs contain one or more variable or constant immunoglobulin domains, but lack other types of domains. Members of the simple-type subgroup include Schwann cell myelin protein (SMP), limbic system-associated membrane protein (LAMP), opiate-binding cell-adhesion molecule (OBCAM), and myelin-associated glycoprotein (MAG). The complex-type NCAPs contain fibronectin type III domains in addition to the immunoglobulin domains. The complex-type subgroup includes neural cell-adhesion molecule (NCAM), axonin-1, F11, Bravo, and L1. Mixed-type NCAPs contain a combination of immunoglobulin domains and other motifs such as tyrosine kinase and epidermal growth factor-like domains. This subgroup includes Trk receptors of nerve growth factors such as nerve growth factor (NGF) and neurotrophin 4 (NT4), Neu differentiation factors such as glial growth factor II (GGFII) and acetylcholine receptor-inducing factor (ARIA), and the semaphorin/collapsin family such as semaphorin B and collapsin.

Semaphorins are a large group of axonal guidance molecules consisting of at least 30 different members and are found in vertebrates, invertebrates, and even certain viruses. All semaphorins contain the sema domain which is approximately 500 amino acids in length. Neuropilin, a semaphorin receptor, has been shown to promote neurite outgrowth *in vitro*. The extracellular region of neuropilins consists of three different domains: CUB, discoidin, and MAM domains. The CUB and the MAM motifs of neuropilin have been proposed to have roles in protein-protein interactions and are suggested to be involved in the binding of semaphorins through the sema and the C-terminal domains (reviewed in Raper, J.A. (2000) Curr. Opin. Neurobiol. 10:88-94).

An NCAP subfamily, the NCAP-LON subgroup, includes cell adhesion proteins expressed on distinct subpopulations of brain neurons. Members of the NCAP-LON subgroup possess three immunoglobulin domains and bind to cell membranes through GPI anchors. Kilon (a kindred of

NCAP-LON), for example, is expressed in the brain cerebral cortex and hippocampus (Funatsu, N. et al. (1999) J. Biol. Chem. 274:8224-8230). Immunostaining localizes Kilon to the dendrites and soma of pyramidal neurons. Kilon has three C2 type immunoglobulin-like domains, six predicted glycosylation sites, and a GPI anchor. Expression of Kilon is developmentally regulated. It is expressed at higher levels in adult brain in comparison to embryonic and early postnatal brains. Confocal microscopy shows the presence of Kilon in dendrites of hypothalamic magnocellular neurons secreting neuropeptides, oxytocin or arginine vasopressin (Miyata, S. et al. (2000) J. Comp. Neurol. 424:74-85). Arginine vasopressin regulates body fluid homeostasis, extracellular osmolarity and intravascular volume. Oxytocin induces contractions of uterine smooth muscle during child birth and of myoepithelial cells in mammary glands during lactation. In magnocellular neurons, Kilon is proposed to play roles in the reorganization of dendritic connections during neuropeptide secretion.

The co-ordinated function of effector and accessory cells in the immune system is assisted by adhesion molecules on the cell surface that stabilize interactions between different cell types.

Leukocyte function-associated antigen 1 (LFA-1) is expressed on the surface of all white blood cells and is a receptor for intercellular adhesion molecules (ICAM) 1 and 2 which are members of the immunoglobulin superfamily. The interaction of LFA-1 with ICAMs 1 and 2 provides essential accessory adhesion signals in many immune interactions, including those between T and B lymphocytes and cytotoxic T cells and their targets. In addition, both ICAMs are expressed at low levels on resting vascular endothelium. ICAM-1 is strongly upregulated by cytokine stimulation and plays a key role in the arrest of leukocytes in blood vessels at sites of inflammation and injury. A third ligand for LFA-1 expressed in resting leukocytes is ICAM-3. ICAM-3 is closely related to ICAM-1 and is constitutively expressed on all leukocytes. It consists of five immunoglobulin domains and binds LFA-1 through its two N-terminal domains (Fawcett, J. et al. (1992) Nature 360:481-484).

Cell adhesion proteins also include some members of the proline-rich proteins (PRPs). PRPs are defined by a high frequency of proline, ranging from 20-50% of the total amino acid content. Some PRPs have short domains which are rich in proline. These proline-rich regions are associated with protein-protein interactions. One family of PRPs are the proline-rich synapse-associated proteins (ProSAPs) which have been shown to bind to members of the postsynaptic density (PSD) protein family and subtypes of the somatostatin receptor (Yao, I. et al. (1999) J. Biol. Chem. 274: 27463-27466; Zitzer, H. et al. (1999) J. Biol. Chem. 274:32997-33001). Members of the ProSAP family contain six to seven ankyrin repeats at the N-terminus, followed by an SH3 domain, a PDZ domain, and seven proline-rich regions and a SAM domain at the C terminus. Several groups of ProSAPs are important structural constituents of synaptic structures in human brain (Zitzer et al., *supra*). Another member of the PRP family is the HLA-B-associated transcript 2 protein (BAT2)

which is rich in proline and includes short tracts of polyproline, polyglycine, and charged amino acids. BAT2 also contains four RGD (Arg-Gly-Asp) motifs typical of integrins (Banerji, J. et al. (1990) Proc. Natl. Acad. Sci. USA 87:2374-2378).

Toposome is a cell-adhesion glycoprotein isolated from mesenchyme-blastula embryos.

- 5 Toposome precursors including vitellogenin promote cell adhesion of dissociated blastula cells.

There are additional specific domains characteristic of cell adhesion proteins. One such domain is the MAM domain, a domain of about 170 amino acids found in the extracellular region of diverse proteins. These proteins all share a receptor-like architecture comprising a signal peptide, followed by a large N-terminal extracellular domain, a transmembrane region, and an intracellular domain (PROSITE document PDOC00604 MAM domain signature and profile). MAM domain proteins include zonadhesin, a sperm-specific membrane protein that binds to the zona pellucida of the egg; neuropilin, a cell adhesion molecule that functions during the formation of certain neuronal circuits, and *Xenopus laevis* thyroid hormone induced protein B, which contains four MAM domains and is involved in metamorphosis (Brown, D.D. et al. (1996) Proc. Natl. Acad. Sci. USA 93:1924-1929).

The WSC domain was originally found in the yeast WSC (cell-wall integrity and stress response component) proteins which act as sensors of environmental stress. The WSC domains are extracellular and are thought to possess a carbohydrate binding role (Ponting, C.P. et al. (1999) Curr. Biol. 9:S1-S2). A WSC domain has recently been identified in polycystin-1, a human plasma membrane protein. Mutations in polycystin-1 are the cause of the commonest form of autosomal dominant polycystic kidney disease (Ponting, C.P. et al. (1999) Curr. Biol. 9:R585-R588).

Leucine rich repeats (LRR) are short motifs found in numerous proteins from a wide range of species. LRR motifs are of variable length, most commonly 20-29 amino acids, and multiple repeats are typically present in tandem. LRR motifs are important for protein/protein interactions and cell adhesion, and LRR proteins are involved in cell/cell interactions, morphogenesis, and development (Kobe, B. and J. Deisenhofer (1995) Curr. Opin. Struct. Biol. 5:409-416). The human ISLR (immunoglobulin superfamily containing leucine-rich repeat) protein contains a C2-type immunoglobulin domain as well as LRR motifs. The ISLR gene is linked to the critical region for Bardet-Biedl syndrome, a developmental disorder of which the most common feature is retinal dystrophy (Nagasawa, A. et al. (1999) Genomics 61:37-43).

The sterile alpha motif (SAM) domain is a conserved protein binding domain, approximately 70 amino acids in length, and is involved in the regulation of many developmental processes in eukaryotes. The SAM domain can potentially function as a protein interaction module through its ability to form homo- or hetero-oligomers with other SAM domains (Schultz, J. et al. (1997) Protein Sci. 6:249-253).

Extracellular Matrix Proteins

The extracellular matrix (ECM) is a complex network of glycoproteins, polysaccharides, proteoglycans, and other macromolecules that are secreted from the cell into the extracellular space. The ECM remains in close association with the cell surface and provides a supportive meshwork that profoundly influences cell shape, motility, strength, flexibility, and adhesion. In fact, adhesion of a cell to its surrounding matrix is required for cell survival except in the case of metastatic tumor cells, which have overcome the need for cell-ECM anchorage. This phenomenon suggests that the ECM plays a critical role in the molecular mechanisms of growth control and metastasis. (Reviewed in Ruoslahti, E. (1996) *Sci. Am.* 275:72-77.) Furthermore, the ECM determines the structure and physical properties of connective tissue and is particularly important for morphogenesis and other processes associated with embryonic development and pattern formation.

The collagens comprise a family of ECM proteins that provide structure to bone, teeth, skin, ligaments, tendons, cartilage, blood vessels, and basement membranes. Multiple collagen proteins have been identified. Three collagen molecules fold together in a triple helix stabilized by interchain disulfide bonds. Bundles of these triple helices then associate to form fibrils. Collagen primary structure consists of hundreds of (Gly-X-Y) repeats where about a third of the X and Y residues are Pro. Glycines are crucial to helix formation as the bulkier amino acid sidechains cannot fold into the triple helical conformation. Because of these strict sequence requirements, mutations in collagen genes have severe consequences. Osteogenesis imperfecta patients have brittle bones that fracture easily; in severe cases patients die *in utero* or at birth. Ehlers-Danlos syndrome patients have hyperelastic skin, hypermobile joints, and susceptibility to aortic and intestinal rupture. Chondrodysplasia patients have short stature and ocular disorders. Alport syndrome patients have hematuria, sensorineural deafness, and eye lens deformation. (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc., New York, NY, pp. 2105-2117; and Creighton, T.E. (1984) Proteins, Structures and Molecular Principles, W.H. Freeman and Company, New York, NY, pp. 191-197.)

Elastin and related proteins confer elasticity to tissues such as skin, blood vessels, and lungs. Elastin is a highly hydrophobic protein of about 750 amino acids that is rich in proline and glycine residues. Elastin molecules are highly cross-linked, forming an extensive extracellular network of fibers and sheets. Elastin fibers are surrounded by a sheath of microfibrils which are composed of a number of glycoproteins, including fibrillin. Mutations in the gene encoding fibrillin are responsible for Marfan's syndrome, a genetic disorder characterized by defects in connective tissue. In severe cases, the aortas of afflicted individuals are prone to rupture. (Reviewed in Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, New York, NY, pp. 984-986.) The fibulin proteins connect elastic fibers and are thought to promote the formation and stabilization of the fiber.

Members of the fibulin family contain epidermal growth factor-like motifs as well as an RGD cell attachment sequence (Midwood, K.S. and J.E. Schwarzbauer (2002) *Current Biology* 12:R279-R281).

Fibronectin is a large ECM glycoprotein found in all vertebrates. Fibronectin exists as a dimer of two subunits, each containing about 2,500 amino acids. Each subunit folds into a rod-like structure containing multiple domains. The domains each contain multiple repeated modules, the most common of which is the type III fibronectin repeat. The type III fibronectin repeat is about 90 amino acids in length and is also found in other ECM proteins and in some plasma membrane and cytoplasmic proteins. Furthermore, some type III fibronectin repeats contain a characteristic tripeptide consisting of Arginine-Glycine-Aspartic acid (RGD). The RGD sequence is recognized by the integrin family of cell surface receptors and is also found in other ECM proteins. Disruption of both copies of the gene encoding fibronectin causes early embryonic lethality in mice. The mutant embryos display extensive morphological defects, including defects in the formation of the notochord, somites, heart, blood vessels, neural tube, and extraembryonic structures. (Reviewed in Alberts et al., *supra*, pp. 986-987.)

Laminin is a major glycoprotein component of the basal lamina which underlies and supports epithelial cell sheets. Laminin is one of the first ECM proteins synthesized in the developing embryo. Laminin is an 850 kilodalton protein composed of three polypeptide chains joined in the shape of a cross by disulfide bonds. Laminin is especially important for angiogenesis and, in particular, for guiding the formation of capillaries. (Reviewed in Alberts et al., *supra*, pp. 990-991.)

Many proteinaceous ECM components are proteoglycans. Proteoglycans are composed of unbranched polysaccharide chains (glycosaminoglycans) attached to protein cores. Common proteoglycans include aggrecan, betaglycan, decorin, perlecan, serglycin, and syndecan-1. Some of these molecules not only provide mechanical support, but also bind to extracellular signaling molecules, such as fibroblast growth factor and transforming growth factor β , suggesting a role for proteoglycans in cell-cell communication. (Reviewed in Alberts et al., *supra*, pp. 973-978.) Likewise, the glycoproteins tenascin-C and tenascin-R are expressed in developing and lesioned neural tissue and provide stimulatory and anti-adhesive (inhibitory) properties, respectively, for axonal growth (Faissner, A. (1997) *Cell Tissue Res.* 290:331-341).

Dentin phosphoryn (DPP) is a major component of the dentin ECM. DPP is a proteoglycan that is synthesized and expressed by odontoblasts (Gu, K. et al. (1998) *Eur. J. Oral Sci.* 106:1043-1047). DPP is believed to nucleate or modulate the formation of hydroxyapatite crystals.

Mucins are highly glycosylated glycoproteins that are the major structural component of the mucus gel. The physiological functions of mucins are cytoprotection, mechanical protection, maintenance of viscosity in secretions, and cellular recognition. MUC6 is a human gastric mucin that is also found in gall bladder, pancreas, seminal vesicles, and female reproductive tract (Toribara,

N.W. et al. (1997) J. Biol. Chem. 272:16398-16403). The MUC6 gene has been mapped to human chromosome 11 (Toribara, N.W. et al. (1993) J. Biol. Chem. 268:5879-5885). Hemomucin is a novel *Drosophila* surface mucin that may be involved in the induction of antibacterial effector molecules (Theopold, U. et al. (1996) J. Biol. Chem. 271:12708-12715).

5 Olfactomedin was originally identified as the major component of the mucus layer surrounding the chemosensory dendrites of olfactory neurons. Olfactomedin-related proteins are secreted glycoproteins with conserved C-terminal motifs. The TIGR/myocilin protein, an olfactomedin-related protein expressed in the eye, is associated with the pathogenesis of glaucoma (Kulkarni, N.H. et al. (2000) Genet. Res. 76:41-50).

10 Ankyrin (ANK) repeats mediate protein-protein interactions associated with diverse intracellular functions. ANK repeats are composed of about 33 amino acids that form a helix-turn-helix core preceded by a protruding "tip." These tips are of variable sequence and may play a role in protein-protein interactions. The helix-turn-helix region of the ANK repeats stack on top of one another and are stabilized by hydrophobic interactions (Yang, Y. et al. (1998) Structure 6:619-626).

15 Sushi repeats, also called short consensus repeats (SCR), are found in a number of proteins that share the common feature of binding to other proteins. For example, in the C-terminal domain of versican, the sushi domain is important for heparin binding. Sushi domains contain basic amino acid residues, which may play a role in binding (Oleszewski, M. et al. (2000) J. Biol. Chem. 275:34478-34485).

20 Link, or X-link, modules are hyaluronan-binding domains found in proteins involved in the assembly of extracellular matrix, cell adhesion, and migration. The Link module superfamily includes CD44, cartilage link protein, and aggrecan. This family also includes BEHAB (brain enriched hyaluronan-binding)/brevican, a component of the brain ECM that is dramatically upregulated in human gliomas, and appears to play a role in determining the invasive potential of
25 brain tumor cells (Gary, S.C. et al. (1998) Curr. Opin. Neurobiol. 8:576-581). There is close similarity between the Link module and the C-type lectin domain, with the predicted hyaluronan-binding site at an analogous position to the carbohydrate-binding pocket in E-selectin (Kohda, D. et al. (1996) Cell 86:767-775).

 Multidomain or mosaic proteins play an important role in the diverse functions of the
30 extracellular matrix (Engel, J. et al. (1994) Development (Camb.):S35-S42). ECM proteins are frequently characterized by the presence of one or more domains which may contain a number of potential intracellular disulfide bridge motifs. For example, domains which match the epidermal growth factor (EGF) tandem repeat consensus are present within several known extracellular proteins that promote cell growth, development, and cell signaling. This signature sequence is about forty
35 amino acid residues in length and includes six conserved cysteine residues, and a calcium-binding site

near the N-terminus of the signature sequence. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines vary in length (Davis, C.G. (1990) *New Biol.* 5:410-419). Post-translational hydroxylation of aspartic acid or asparagine residues has been associated with EGF-like domains in several proteins (Prosite PDOC00010).

A number of proteins that contain calcium-binding EGF-like domain signature sequences are involved in growth and differentiation. Examples include bone morphogenic protein 1, which induces the formation of cartilage and bone; crumbs, which is a *Drosophila* epithelial development protein; Notch and a number of its homologs, which are involved in neural growth and differentiation, and transforming growth factor beta-1 binding protein (Expasy PROSITE document PDOC00913; Soler, C. and G. Carpenter, in Nicola, N.A. (1994) *The Cytokine Facts Book*, Oxford University Press, Oxford, UK, pp. 193-197). EGF-like domains mediate protein-protein interactions for a variety of proteins. For example, EGF-like domains in the ECM glycoprotein fibulin-1 have been shown to mediate both self-association and binding to fibronectin (Tran, H. et al. (1997) *J. Biol. Chem.* 272:22600-22606). Point mutations in the EGF-like domains of ECM proteins have been identified as the cause of human disorders such as Marfan syndrome and pseudochoondroplasia (Maurer, P. et al. (1996) *Curr. Opin. Cell Biol.* 8:609-617).

The CUB domain is an extracellular domain of approximately 110 amino acid residues found mostly in developmentally regulated proteins. The CUB domain contains four conserved cysteine residues and is predicted to have a structure similar to that of immunoglobulins. Vertebrate bone morphogenic protein 1, which induces cartilage and bone formation, and fibropellins I and III from sea urchin, which form the apical lamina component of the ECM, are examples of proteins that contain both CUB and EGF domains (PROSITE PDOC00908).

Other ECM proteins are members of the type A domain of von Willebrand factor (vWFA)-like module superfamily, a diverse group of proteins with a module sharing high sequence similarity. The vWFA-like module is found not only in plasma proteins but also in plasma membrane and ECM proteins (Colombatti, A. and P. Bonaldo (1991) *Blood* 77:2305-2315). Crystal structure analysis of an integrin vWFA-like module shows a classic "Rossmann" fold and suggests a metal ion-dependent adhesion site for binding protein ligands (Lee, J.-O. et al. (1995) *Cell* 80:631-638). This family includes the protein matrilin-2, an extracellular matrix protein that is expressed in a broad range of mammalian tissues and organs. Matrilin-2 is thought to play a role in ECM assembly by bridging collagen fibrils and the aggrecan network (Deak, F. et al. (1997) *J. Biol. Chem.* 272:9268-9274).

The thrombospondins are multimeric, calcium-binding extracellular glycoproteins found widely in the embryonic extracellular matrix. These proteins are expressed in the developing nervous system or at specific sites in the adult nervous system after injury. Thrombospondins contain

multiple EGF-type repeats, as well as a motif known as the thrombospondin type 1 repeat (TSR). The TSR is approximately 60 amino acids in length and contains six conserved cysteine residues. Motifs within TSR domains are involved in mediating cell adhesion through binding to proteoglycans and sulfated glycolipids. Thrombospondin-1 inhibits angiogenesis and modulates endothelial cell adhesion, motility, and growth. TSR domains are found in a diverse group of other proteins, most of which are expressed in the developing nervous system and have potential roles in the guidance of cell and growth cone migration. Proteins that contain TSRs include the F-spondin gene family, the semaphorin 5 family, UNC-5, and SCO-spondin. The TSR superfamily includes the ADAMTS proteins which contain an ADAM (A Disintegrin and Metalloproteinase) domain as well as one or more TSRs. The ADAMTS proteins have roles in regulating the turnover of cartilage matrix, regulation of blood vessel growth, and possibly development of the nervous system. (Reviewed in Adams, J.C. and R.P. Tucker (2000) Dev. Dyn. 218:280-299.)

Fibrinogen, the principle protein of vertebrate blood clotting, is a hexamer consisting of two sets of three different chains (alpha, beta, and gamma). The C-terminal domain of the beta and gamma chains comprises about 270 amino acid residues and contains four cysteines involved in two disulfide bonds. This domain has also been found in mammalian tenascin-X, an ECM protein that appears to be involved in cell adhesion (Prosite PDOC00445).

Osteopontin (OPN) is a phosphorylated glycoprotein that contains a functional Gly-Arg-Gly-Asp-Ser (GRGDS) cell-binding sequence. OPN is expressed in several glioma cell lines. Expression of OPN mRNA was found to be high in malignant astrocytomas but low in benign astrocytomas and non-neoplastic tissue (Saitoh, Y. et al. (1995) Lab. Invest. 72:55-63). One glioma-derived cell line which does not express significant amounts of OPN mRNA, was induced in a dose-dependent manner by the tumor-promoting and PKC-activating phorbol ester, TPA, to over-express OPN mRNA in a PKC-dependent manner. Treatment of these cells with Ca²⁺ ionophore (A23187) completely inhibited TPA-mediated induction of OPN while treatment with the intracellular Ca²⁺ antagonist TMB-8 had no significant effect (Tucker, M.A. et al. (1998) Anticancer Res. 18:807-812).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the

expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

Array technology can provide a simple way to explore the expression profile of a large number of related or unrelated genes. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder. The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with lung cancer may be compared with the levels and sequences expressed in normal tissue.

Lung Cancer

Lung cancer is the leading cause of cancer death in the United States, affecting more than 100,000 men and 50,000 women each year. Nearly 90% of the patients diagnosed with lung cancer are cigarette smokers. Tobacco smoke contains thousands of noxious substances that induce carcinogen metabolizing enzymes and covalent DNA adduct formation in the exposed bronchial epithelium. In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. This adversely affects the overall five-year survival rate which is 37% for squamous carcinoma, 27% for adenocarcinoma and large cell carcinoma, and less than 1% for small cell carcinomas. Earlier diagnosis and an systematic approach to identification, staging, and treatment could positively affect patient outcome (DeVita et al. (1997) Cancer: Principles and Practice of Oncology, Lippincott-Raven, Philadelphia PA) and Fauci et al. (1998) Harrison's Principals of Internal Medicine, McGraw Hill, New York, NY). In nearly 80% of patients diagnosed with lung cancer, metastasis has already occurred. Most commonly lung cancers metastasize to pleura, brain, bone, pericardium, and liver. The decision to treat with surgery, radiation therapy, or chemotherapy is made on the basis of tumor histology, response to growth factors or hormones, and sensitivity to inhibitors or drugs. With current treatments, most patients die within one year of diagnosis. Earlier diagnosis and a systematic approach to identification, staging, and treatment of lung cancer could positively affect patient outcome.

Lung cancers progress through a series of morphologically distinct stages from hyperplasia to invasive carcinoma. Malignant lung cancers are divided into two groups comprising four

histopathological classes. The Non Small Cell Lung Carcinoma (NSCLC) group includes squamous cell carcinomas, adenocarcinomas, and large cell carcinomas and accounts for about 70% of all lung cancer cases. Adenocarcinomas typically arise in the peripheral airways and often form mucin secreting glands. Squamous cell carcinomas typically arise in proximal airways. The histogenesis of squamous cell carcinomas may be related to chronic inflammation and injury to the bronchial epithelium, leading to squamous metaplasia. The Small Cell Lung Carcinoma (SCLC) group accounts for about 20% of lung cancer cases. SCLCs typically arise in proximal airways and exhibit a number of paraneoplastic syndromes including inappropriate production of adrenocorticotropin and anti-diuretic hormone.

Lung cancer cells accumulate numerous genetic lesions, many of which are associated with cytologically visible chromosomal aberrations. The high frequency of chromosomal deletions associated with lung cancer may reflect the role of multiple tumor suppressor loci in the etiology of this disease. Deletion of the short arm of chromosome 3 is found in over 90% of cases and represents one of the earliest genetic lesions leading to lung cancer. Several studies report deletions of regions of chromosome 11 in NSCLC (Bepler, G. and Garcia-Blanco, M.A. (1994) PNAS 91:5513-7; Iizuka, M., et al. (1995) Genes, Chromosomes & Cancer 13:40-46; Rasio, D. (1995) Cancer Research 55:3988-91). Deletions in other chromosome arms such as 3p, 9p and 17p are also common. Other frequently observed genetic lesions include overexpression of telomerase, activation of oncogenes such as K-ras and c-myc, and inactivation of tumor suppressor genes such as RB, CDKN2, p53 and p16 (Toomey, D. et al. (2001) Cancer 92:2648-57; Zajac-Kaye M. (2001) Lung Cancer 34:S43-6; Wright, G. et al. (2000) Current Opinion in Oncology 12:143-8; Kohno, T. and Yokota, J. (1999) Carcinogenesis 20:1403-10).

Genes differentially regulated in lung cancer have been identified by a variety of methods. Using mRNA differential display technology, Manda et al. (1999; Genomics 51:5-14) identified five genes differentially expressed in lung cancer cell lines compared to normal bronchial epithelial cells. Among the known genes, pulmonary surfactant apoprotein A and alpha 2 macroglobulin were down regulated whereas nm23H1 was upregulated. Petersen et al. (2000; Int J. Cancer, 86:512-517) used suppression subtractive hybridization to identify 552 clones differentially expressed in lung tumor derived cell lines, 205 of which represented known genes. Among the known genes, thrombospondin-1, fibronectin, intercellular adhesion molecule 1, and cytokeratins 6 and 18 were previously observed to be differentially expressed in lung cancers. Wang et al. (2000; Oncogene 19:1519-1528) used a combination of microarray analysis and subtractive hybridization to identify 17 genes differentially overexpressed in squamous cell carcinoma compared with normal lung epithelium. Among the known genes they identified were keratin isoform 6, KOC, SPRC, IGFb2, connexin 26, plakophilin 1 and cytokeratin 13.

Senescence

Senescence is a normal mechanism of tumor suppression, a homeostatic device that evolved to limit cell proliferation and protect the organism against cancer. The proliferative lifespan of most normal human cells, even in ideal growth conditions, is limited by intrinsic inhibitory signals that induce cell cycle arrest after a preset number of cell divisions. This process of "replicative senescence" is activated in many cell types by the progressive deletion of the specialized ends of chromosomes — telomeres — which act as molecular "clocks". A number of molecular changes observed in replicative senescent cells occur in somatic cells during the process of aging. Genetic studies on replicative senescence indicate the control of tumor suppression mechanisms. Despite the significance of replicative senescence in aging and cancer, little is known about the central cause of the complex changes observed in replicative senescent cells. Despite the protection from cancer conveyed by cellular senescence and other mechanisms that suppress tumorigenesis, the development of cancer is almost inevitable as mammalian organisms age. Certainly, aging predisposes cells to accumulate mutations, several of which may eventually cause malignant transformations, particularly in humans. However, many benign or relatively well-controlled tumors may also harbor many potentially oncogenic mutations, suggesting that the tissue microenvironment can suppress the expression of many malignant phenotypes. Although the idea remains controversial, cellular senescence has also been proposed to contribute to organismal aging. Senescent cells have recently been shown to accumulate with age in human tissues. One possibility is that the tissue microenvironment is disrupted by the accumulation of dysfunctional senescent cells. Thus, mutation accumulation may synergize with the accumulation of senescent cells, leading to the increased risk for developing cancer that is a hallmark of mammalian aging.

The molecular basis of cancer development has been studied extensively. Many oncogenes immortalize cells as a first step toward tumor formation. Normally, most eukaryotic cells, after a certain number of divisions, enter a state of senescence in which cells remain viable and metabolically active but no longer replicate. A number of phenotypic changes, such as increased cell size and pH-dependent beta-galactosidase activity, and molecular changes, such as the upregulation of particular genes, occur in senescent cells (Shelton (1999) Curr. Biol. 9:939-945). When senescent cells are exposed to mitogens, a number of genes are upregulated, but the cells do not proliferate. Evidence indicates that senescent cells accumulate with age *in vivo*, contributing to the aging of an organism. In addition, senescence suppresses tumorigenesis; and many genes necessary for senescence also function as tumor suppressor genes, such as p53 and the retinoblastoma susceptibility gene. Tumor cells must overcome this checkpoint for proliferation. Most tumors contain cells that have surpassed their replicative limit, i.e. they are immortalized.

A variety of challenges, such as oxidative stress, radiation, activated oncoproteins, and cell

cycle inhibitors, induce a senescent phenotype, indicating that senescence is influenced by a number of proliferative and anti-proliferative signals (Shelton, *supra*). Senescence is correlated with the progressive shortening of telomeres that occurs with each cell division. Expression of the catalytic component of telomerase in cells prevents telomere shortening and immortalizes cells such as fibroblasts and epithelial cells, but not other types of cells, such as CD8+ T cells (Migliaccio et al.(2000) J. Immunol. 165:4978-4984). Thus, senescence is controlled by telomere shortening as well as other mechanisms depending on the type of cell.

A number of genes that are differentially expressed between senescent and presenescent cells have been identified as part of ongoing studies to understand the role of senescence in aging and tumorigenesis. Most senescent cells are growth arrested in the G1 stage of the cell cycle. While expression of many cell cycle genes is similar in senescent and presenescent cells (Cristofalo (1992) Ann. NY Acad. Sci. 663:187-194), expression of others genes such as cyclin-dependent kinases p21 and p16, which inhibit proliferation, and cyclins D1 and E is elevated in senescent cells. Other genes that are not directly involved in the cell cycle are also upregulated such as extracellular matrix proteins--fibronectin, procollagen, and osteonectin; and proteases such as collagenase, stromelysin, and cathepsin B (Chen (2000) Ann. NY Acad. Sci. 908:111-125). Genes underexpressed in senescent cells include those that encode heat shock proteins, c-fos, and cdc-2 (Chen, *supra*). The further identification of genes controlling the lifespan of cells will be facilitated by array technology.

Steroid Hormones

Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus. Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catecholamines in the central nervous system, and reduce inflammation. The principal mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex

hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

5 Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and auto-immune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and
10 for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release
15 of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They
20 inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions.

Beclomethasone is a synthetic glucocorticoid used to treat steroid-dependent asthma, to relieve
25 symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central

30 nervous system, dexamethasone is usually the treatment of choice to control cerebral edema.

Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone
35 and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus

erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions.

The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A₂ inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of β -adrenergic receptors on leukocytes, and decreased arachidonic acid metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can be retarded by administration of corticosteroids.

Breast Cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (K. Gish (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou CM et al. (2000) Nature 406:747-752).

Breast cancer is a genetic disease commonly caused by mutations in cellular disease. Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, supra). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to noninherited mutations that occur in breast epithelial cells.

A good deal is already known about the expression of specific genes associated with breast cancer. For example, the relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie et al., supra, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related

to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, SS et al. (1994) Am J Clin Pathol 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix Gla protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou Z et al. (1998) Int J Cancer 78:95-99; Chen, L et al. (1990) Oncogene 5:1391-1395; Ulrix W et al. (1999) FEBS Lett 455:23-26; Sager, R et al. (1996) Curr Top Microbiol Immunol 213:51-64; and Lee, SW et al. (1992) Proc Natl Acad Sci USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba II et al. (1998) Clin Cancer Res 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Colon Cancer

Colorectal cancer is the fourth most common cancer and the second most common cause of cancer death in the United States with approximately 130,000 new cases and 55,000 deaths per year. Colon and rectal cancers share many environmental risk factors and both are found in individuals with specific genetic syndromes. (See Potter, JD (1999) J Natl Cancer Institute 91:916-932 for a review of colorectal cancer.) Colon cancer is the only cancer that occurs with approximately equal frequency in men and women, and the five-year survival rate following diagnosis of colon cancer is around 55% in the United States (Ries et al. (1990) National Institutes of Health, DHHS Publ No. (NIH)90-2789).

Colon cancer is causally related to both genes and the environment. Several molecular pathways have been linked to the development of colon cancer, and the expression of key genes in any of these pathways may be lost by inherited or acquired mutation or by hypermethylation. There is a particular need to identify genes for which changes in expression may provide an early indicator of colon cancer or a predisposition for the development of colon cancer.

For example, it is well known that abnormal patterns of DNA methylation occur consistently

in human tumors and include, simultaneously, widespread genomic hypomethylation and localized areas of increased methylation. In colon cancer in particular, it has been found that these changes occur early in tumor progression such as in premalignant polyps that precede colon cancer. Indeed, DNA methyltransferase, the enzyme that performs DNA methylation, is significantly increased in histologically normal mucosa from patients with colon cancer or the benign polyps that precede cancer, and this increase continues during the progression of colonic neoplasms (Wafik, S et al. (1991) Proc Natl Acad Sci USA 88:3470-3474). Increased DNA methylation occurs in G+C rich areas of genomic DNA termed "CpG islands" that are important for maintenance of an "open" transcriptional conformation around genes, and that hypermethylation of these regions results in a "closed" conformation that silences gene transcription. It has been suggested that the silencing or downregulation of differentiation genes by such abnormal methylation of CpG islands may prevent differentiation in immortalized cells (Antequera, F. et al. (1990) Cell 62:503-514).

Familial Adenomatous Polyposis (FAP) is a rare autosomal dominant syndrome that precedes colon cancer and is caused by an inherited mutation in the adenomatous polyposis coli (APC) gene. FAP is characterized by the early development of multiple colorectal adenomas that progress to cancer at a mean age of 44 years. The APC gene is a part of the APC- β -catenin-Tcf (T-cell factor) pathway. Impairment of this pathway results in the loss of orderly replication, adhesion, and migration of colonic epithelial cells that results in the growth of polyps. A series of other genetic changes follow activation of the APC- β -catenin-Tcf pathway and accompanies the transition from normal colonic mucosa to metastatic carcinoma. These changes include mutation of the K-Ras proto-oncogene, changes in methylation patterns, and mutation or loss of the tumor suppressor genes p53 and Smad4/DPC4. While the inheritance of a mutated APC gene is a rare event, the loss or mutation of APC and the consequent effects on the APC- β -catenin-Tcf pathway is believed to be central to the majority of colon cancers in the general population.

Hereditary nonpolyposis Colorectal Cancer (HNPCC) is another inherited autosomal dominant syndrome with a less well defined phenotype than FAP. HNPCC, which accounts for about 2% of colorectal cancer cases, is distinguished by the tendency to early onset of cancer and the development of other cancers, particularly those involving the endometrium, urinary tract, stomach and biliary system. HNPCC results from the mutation of one or more genes in the DNA mismatch repair (MMR) pathway. Mutations in two human MMR genes, MSH2 and MLH1, are found in a large majority of HNPCC families identified to date. The DNA MMR pathway identifies and repairs errors that result from the activity of DNA polymerase during replication. Furthermore, loss of MMR activity contributes to cancer progression through accumulation of other gene mutations and deletions, such as loss of the BAX gene which controls apoptosis, and the TGF β receptor II gene

which controls cell growth. Because of the potential for irreparable damage to DNA in an individual with a DNA MMR defect, progression to carcinoma is more rapid than usual.

Although ulcerative colitis is a minor contributor to colon cancer, affected individuals have about a 20-fold increase in risk for developing cancer. Progression is characterized by loss of the p53 gene which may occur early, appearing even in histologically normal tissue. The progression of the disease from ulcerative colitis to dysplasia/carcinoma without an intermediate polyp state suggests a high degree of mutagenic activity resulting from the exposure of proliferating cells in the colonic mucosa to the colonic contents.

Almost all colon cancers arise from cells in which the estrogen receptor (ER) gene has been silenced. The silencing of ER gene transcription is age related and linked to hypermethylation of the ER gene (Issa, J-P J et al. (1994) Nature Genetics 7:536-540). Introduction of an exogenous ER gene into cultured colon carcinoma cells results in marked growth suppression. The connection between loss of the ER protein in colonic epithelial cells and the consequent development of cancer has not been established.

Clearly there are a number of genetic alterations associated with colon cancer and with the development and progression of the disease, particularly the downregulation or deletion of genes, that potentially provide early indicators of cancer development, and which may also be used to monitor disease progression or provide possible therapeutic targets. The specific genes affected in a given case of colon cancer depend on the molecular progression of the disease. Identification of additional genes associated with colon cancer and the precancerous state would provide more reliable diagnostic patterns associated with the development and progression of the disease.

Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rates for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. The molecular events that lead to ovarian cancer are poorly understood. Some of the known aberrations include mutation of p53 and microsatellite instability. Since gene expression patterns are likely to vary when normal ovary is compared to ovarian tumors, examination of gene expression in these tissues to identify possible markers for ovarian cancer is particularly relevant to improving diagnosis, prognosis, and treatment of this disease.

Prostate Cancer

Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression

of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for examining which genes are tissue specific, carrying out housekeeping functions, parts of a signaling cascade, or specifically related to a particular genetic predisposition, condition, disease, or disorder.

5 The potential application of gene expression profiling is particularly relevant to improving diagnosis, prognosis, and treatment of disease. For example, both the levels and sequences expressed in tissues from subjects with prostate or with breast cancer may be compared with the levels and sequences expressed in normal tissue.

10 Prostate cancer is a common malignancy in men over the age of 50, and the incidence increases with age. In the US, there are approximately 132,000 newly diagnosed cases of prostate cancer and more than 33,000 deaths from the disorder each year.

15 Once cancer cells arise in the prostate, they are stimulated by testosterone to a more rapid growth. Thus, removal of the testes can indirectly reduce both rapid growth and metastasis of the cancer. Over 95 percent of prostatic cancers are adenocarcinomas which originate in the prostatic acini. The remaining 5 percent are divided between squamous cell and transitional cell carcinomas, both of which arise in the prostatic ducts or other parts of the prostate gland.

20 As with most cancers, prostate cancer develops through a multistage progression ultimately resulting in an aggressive, metastatic phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells that become hyperplastic and evolve into early-stage tumors. The early-stage tumors are localized in the prostate but eventually may metastasize, particularly to the bone, brain or lung. About 80% of these tumors remain responsive to androgen treatment, an important hormone controlling the growth of prostate epithelial cells. However, in its most advanced state, cancer growth becomes androgen-independent and there is currently no known treatment for this condition.

25 A primary diagnostic marker for prostate cancer is prostate specific antigen (PSA). PSA is a tissue-specific serine protease almost exclusively produced by prostatic epithelial cells. The quantity of PSA correlates with the number and volume of the prostatic epithelial cells, and consequently, the levels of PSA are an excellent indicator of abnormal prostate growth. Men with prostate cancer exhibit an early linear increase in PSA levels followed by an exponential increase prior to diagnosis. 30 However, since PSA levels are also influenced by factors such as inflammation, androgen and other growth factors, some scientists maintain that changes in PSA levels are not useful in detecting individual cases of prostate cancer.

 Current areas of cancer research provide additional prospects for markers as well as potential therapeutic targets for prostate cancer. Several growth factors have been shown to play a critical role

in tumor development, growth, and progression. The growth factors Epidermal Growth Factor (EGF), Fibroblast Growth Factor (FGF), and Tumor Growth Factor alpha (TGF α) are important in the growth of normal as well as hyperproliferative prostate epithelial cells, particularly at early stages of tumor development and progression, and affect signaling pathways in these cells in various ways (Lin J et al. (1999) Cancer Res. 59:2891-2897; Putz T et al. (1999) Cancer Res 59:227-233). The TGF- β family of growth factors are generally expressed at increased levels in human cancers and the high expression levels in many cases correlates with advanced stages of malignancy and poor survival (Gold LI (1999) Crit Rev Oncog 10:303-360). Finally, there are human cell lines representing both the androgen-dependent stage of prostate cancer (LNCap) as well as the androgen-independent, hormone refractory stage of the disease (PC3 and DU-145) that have proved useful in studying gene expression patterns associated with the progression of prostate cancer, and the effects of cell treatments on these expressed genes (Chung TD (1999) Prostate 15:199-207).

Endometrial Cancer

Endometrial cancer is the most common gynecologic cancer. Approximately 90% of endometrial cancers are epithelial in origin, and 90% of these cancers are classified as endometrial adenocarcinomas. Estrogen appears to act as a tumor promoter in endometrial tissue. Evidence indicates that p53 and Ki-ras are mutated in endometrial cancer. However, these mutations occur in a small percentage of cases and do not appear to be the initiating events in the disease. In addition, most chromosomes contain regions of allelic loss in endometrial cancer, indicating that many genes may be affected in this disease.

Osteosarcoma

Osteosarcoma is the most common malignant bone tumor in children. Approximately 80% of patients present with non-metastatic disease. After the diagnosis is made by an initial biopsy, treatment involves the use of 3–4 courses of neoadjuvant chemotherapy before definitive surgery, followed by post-operative chemotherapy. With currently available treatment regimens, approximately 30–40% of patients with non-metastatic disease relapse after therapy. Currently, there is no prognostic factor that can be used at the time of initial diagnosis to predict which patients will have a high risk of relapse. The only significant prognostic factor predicting the outcome in a patient with non-metastatic osteosarcoma is the histopathologic response of the primary tumor resected at the time of definitive surgery. The degree of necrosis in the primary tumor is a reflection of the tumor response to neoadjuvant chemotherapy. A higher degree of necrosis (good or favorable response) is associated with a lower risk of relapse and a better outcome. Patients with a lower degree of necrosis (poor or unfavorable response) have a much higher risk of relapse and poor outcome even after complete resection of the primary tumor. Unfortunately, poor outcome cannot be altered despite

modification of post-operative chemotherapy to account for the resistance of the primary tumor to neoadjuvant chemotherapy. Thus, there is an urgent need to identify prognostic factors that can be used at the time of diagnosis to recognize the subtypes of osteosarcomas that have various risks of relapse, so that more appropriate chemotherapy can be used at the outset to improve the outcome. We
5 hypothesize that global gene expression analysis is a systematic and unbiased approach to recognize such tumor subtypes.

Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common cause of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and provide care for this
10 complex disease has not been achieved. Molecular characterization of growth and regression of atherosclerotic vascular lesions requires identification of the genes that contribute to features of the lesion including growth, stability, dissolution, rupture and, most lethally, induction of occlusive vessel thrombus. Vascular lesions principally involve the vascular endothelium and the surrounding smooth muscle tissue.

Blood vessel walls are composed of two tissue layers: an endothelial cell (EC) layer which comprises the luminal surface of the vessel, and an underlying vascular smooth muscle cell (VSMC)
15 layer. Through dynamic interactions with each other and with surrounding tissues, the vascular endothelium and smooth muscle tissues maintain vascular tone, control selective permeability of the vascular wall, direct vessel remodeling and angiogenesis, and modulate inflammatory and immune
20 responses.

The inflammatory response is a complex vascular reaction mediated by numerous cytokines, chemokines, growth factors, and other signaling molecules expressed by activated ECs, VSMCs and leukocytes. Inflammation protects the organism during trauma and infection, but can also lead to pathological conditions such as atherosclerosis. The pro-inflammatory cytokines, interleukin (IL)-1
25 and tumor necrosis factor (TNF), are secreted by a small number of activated macrophages or other cells and can set off a cascade of vascular changes, largely through their ability to alter gene expression patterns in ECs and VSMCs. These vascular changes include vasodilation and increased permeability of microvasculature, edema, and leukocyte extravasation and transmigration across the vessel wall. Ultimately, leukocytes, particularly neutrophils and monocytes/macrophages, accumulate
30 in the extravascular space, where they remove injurious agents by phagocytosis and oxidative killing, a process accompanied by release of toxic factors, such as proteases and reactive oxygen species.

IL-1 and TNF induce pro-inflammatory, thrombotic, and anti-apoptotic changes in gene expression by signaling through receptors on the surface of ECs and VSMCs; these receptors activate transcription factors such as NF κ B as well as AP-1, IRF-1, and NF-GMa, leading to alterations in

gene expression. Genes known to be differentially regulated in EC by IL-1 and TNF include E selectin, VCAM-1, ICAM-1, PAF, I κ B α , IAP-1, MCP-1, eotaxin, ENA-78, G-CSF, A20, ICE, and complement C3 component. A key event in inflammation, adhesion and transmigration of blood leukocytes across the vascular endothelium, for example, is mediated by increased expression of E selectin, P selectin, ICAM-1, and VCAM-1 on activated endothelium.

Several investigators have examined changes in vascular cell gene expression associated with various inflammatory diseases or model systems. Examining human umbilical vein endothelial cells (HUVEC) activated by recombinant TNF α or conditioned medium from activated human primary monocytes, Horrevoets *et al.* (1999; Blood 93:3418-3431) identified 106 differentially regulated genes. In a similar approach, deVries *et al.* (2000; JBC 275:23939-23947) identified 40 differentially regulated genes in umbilical cord artery-derived smooth muscle cells activated by conditioned media from cultured macrophages after stimulation with oxidized LDL particles. In both studies, many of the identified genes were already known to be involved in inflammation. Comparing expression profiles from inflammatory diseased tissues, cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synoviocytes, Heller *et al.* (1997; Proc Natl Acad Sci USA 94:2150-2155) identified candidate genes involved in inflammatory responses, including TNF, IL-1 IL-6, IL-8 G-CSF, RANTES, and V-CAM. From this candidate gene set, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase were found to be differentially expressed in rheumatoid arthritis (RA) relative to inflammatory bowel disease (IBD). Further, IL-3, chemokine Gro α , and metalloproteinase matrix metallo-elastase were expressed in both RA and IBD. Most recently, in an analysis of cultured aortic smooth muscle cells treated with TNF α , Haley *et al.* (2000; Circulation 102:2185-2189) found a 20-fold increase in eotaxin, an eosinophil chemotactic factor. The overexpression of eotaxin and its receptor CCR3 in atherosclerotic lesions was confirmed by northern analysis.

Development of atherosclerosis is understood to be induced by the presence of circulating lipoprotein. Lipoproteins, such as the cholesterol-rich low-density lipoprotein (LDL), accumulate in the extracellular space of the vascular intima, and undergo modification. Oxidation of LDL (Ox-LDL) occurs most avidly in the sub-endothelial space where circulating antioxidant defenses are less effective. Mononuclear phagocytes enter the intima, differentiate into macrophages, and ingest modified lipids including Ox-LDL. During Ox-LDL uptake, macrophages produce cytokines (e.g. tumor necrosis factor α (TNF- α) and interleukin-1 (IL-1)) and growth factors (e.g. M-CSF, VEGF, and PDGF-BB) that elicit further cellular events that modulate atherogenesis such as smooth muscle cell proliferation and production of extracellular matrix by vascular endothelium. Additionally, these macrophages may activate genes in endothelium and smooth muscle tissue involved in inflammation

and tissue differentiation, including superoxide dismutase (SOD), IL-8, and ICAM-1.

The vascular endothelium influences not only the three classically interacting components of hemostasis: the vessel, the blood platelets and the clotting and fibrinolytic systems of plasma, but also the natural sequelae: inflammation and tissue repair. Two principal modes of endothelial behavior may be differentiated, best defined as an anti- and a prothrombotic state. Under physiological conditions endothelium mediates vascular dilatation (formation of nitric oxide (NO), PGI₂, adenosine, hyperpolarising factor), prevents platelet adhesion and activation (production of adenosine, NO and PGI₂, removal of ADP), blocks thrombin formation (tissue factor pathway inhibitor, activation of protein C via thrombomodulin, activation of antithrombin III) and mitigates fibrin deposition (t- and scu plasminogen activator production). Adhesion and transmigration of inflammatory leukocytes are attenuated, e.g. by NO and IL-10, and oxygen radicals are efficiently scavenged (urate, NO, glutathione, SOD).

When the endothelium is physically disrupted or functionally perturbed by postischemic reperfusion, acute and chronic inflammation, atherosclerosis, diabetes and chronic arterial hypertension, then completely opposing actions pertain. This prothrombotic, proinflammatory state is characterised by vaso-constriction, platelet and leukocyte activation and adhesion (externalisation, expression and upregulation of, for example, von Willebrand factor, platelet activating factor, P-selectin, ICAM-1, IL-8, MCP-1, and TNF- α), promotion of thrombin formation, coagulation and fibrin deposition at the vascular wall (expression of tissue factor, PAI-1, and phosphatidyl serine) and, in platelet-leukocyte coaggregates, additional inflammatory interactions via attachment of platelet CD40-ligand to endothelial, monocyte and B-cell CD40. Since thrombin formation and inflammatory stimulation set the stage for later tissue repair, complete abolition of such endothelial responses cannot be the goal of clinical interventions aimed at limiting procoagulatory, prothrombotic actions of a dysfunctional vascular endothelium. (See, e.g., Becker et al. (2000) Z. Kardiol. 89:160-167.)

Tumor necrosis factor α is a pleiotropic cytokine that mediates immune regulation and inflammatory responses. TNF- α -related cytokines generate partially overlapping cellular responses, including differentiation, proliferation, nuclear factor- κ B (NF- κ B) activation, and cell death, by triggering the aggregation of receptor monomers (Smith, C.A. et al. (1994) Cell 76:959-962). The cellular responses triggered by TNF- α are initiated through its interaction with distinct cell surface receptors (TNFRs). NF- κ B is a transcription factor with a pivotal role in inducing genes involved in physiological processes as well as in the response to injury and infection. Activation of NF- κ B involves the phosphorylation and subsequent degradation of an inhibitory protein, I κ B, and many of the proximal kinases and adaptor molecules involved in this process have been elucidated.

Additionally, the NF- κ B activation pathway from cell membrane to nucleus for IL-1 and TNF- α is

now understood (Bowie and O'Neill (2000) Biochem. Pharmacol. 59:13-23).

Monocyte chemoattractant protein-1 (MCP-1) is known to play an important role in the pathogenesis of atherosclerosis by inducing monocyte migration. TNF- α treatment of human umbilical vein endothelial cells (HUVECs) increased the cellular secretions of MCP-1 119-fold compared with untreated cells. Troglitazone, an insulin-sensitizing drug, significantly inhibited this TNF- α -induced increase in MCP-1 secretions and decreased mRNA levels (Ohta et al. (2000) Diabetes Res. Clin. Pract. 48:171-176).

Treatment of confluent cultures of vascular smooth muscle cells (SMCs) with TNF- α suppresses the incorporation of [3 H]proline into both collagenase-digestible proteins (CDP) and noncollagenous proteins (NCP). Such suppression by TNF- α is not observed in confluent bovine aortic endothelial cells and human fibroblastic IMR-90 cells. TNF- α decreases the relative proportion of collagen types IV and V suggesting that TNF- α modulates collagen synthesis by SMCs depending on their cell density and therefore may modify formation of atherosclerotic lesions (Hiraga et al. (2000) Life Sci. 66:235-244).

Human coronary artery smooth muscle cells (CASMC) are primary cells isolated from the tunica media (an intermediate muscular layer) of a human coronary artery. Vascular smooth muscle cells are a model of increasing significance in vascular biology. It is now well known that besides their obvious role in the regulation of vascular tone and consequently, oxygen supply to various tissues, their behavior under inflammatory conditions is an important factor in the development of atherosclerosis and restenosis.

Human aortic endothelial cells (HAECs) are primary cells derived from the endothelium of a human aorta. HAECs have been used as an experimental model for investigating *in vitro* the role of the endothelium in human vascular biology. Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation. Thus, vascular tissue genes differentially expressed during treatment of CASMC and HAEC cell cultures with TNF α may reasonably be expected to be markers of the atherosclerotic process.

The primary function of adipose tissue is the ability to store and release fat during periods of feeding and fasting. White adipose tissue is the major energy reserve in periods of fasting, and its reserve is mobilized during energy deprivation. Adipose tissue is one of the primary target tissues for insulin, and adipogenesis and insulin resistance are linked in type II diabetes, non-insulin dependent diabetes mellitus (NIDDM). Cytologically the conversion of a preadipocytes into mature adipocytes is characterized by deposition of fat droplets around the nuclei. The conversion process *in vivo* can be induced by thiazolidinediones (TZDs) and other PPAR γ agonists (Adams et al. (1997) J. Clin.

Invest. 100:3149-3153) which also lead to increased sensitivity to insulin and reduced plasma glucose and blood pressure.

Thiazolidinediones (TZDs) act as agonists for the peroxisome-proliferator-activated receptor gamma (PPAR γ), a member of the nuclear hormone receptor superfamily. TZDs reduce
5 hyperglycemia, hyperinsulinemia, and hypertension, in part by promoting glucose metabolism and inhibiting gluconeogenesis. Roles for PPAR γ and its agonists have been demonstrated in a wide range of pathological conditions including diabetes, obesity, hypertension, atherosclerosis, polycystic ovarian syndrome, and cancers such as breast, prostate, liposarcoma, and colon cancer.

The mechanism by which TZDs and other PPAR γ agonists enhance insulin sensitivity is not
10 fully understood, but may involve the ability of PPAR γ to promote adipogenesis. When ectopically expressed in cultured preadipocytes, PPAR γ is a potent inducer of adipocyte differentiation. TZDs, in combination with insulin and other factors, can also enhance differentiation of human preadipocytes in culture (Adams et al. (1997) J. Clin. Invest. 100:3149-3153). The relative potency of different TZDs in promoting adipogenesis in vitro is proportional to both their insulin sensitizing effects in
15 vivo, and their ability to bind and activate PPAR γ in vitro. Interestingly, adipocytes derived from omental adipose depots are refractory to the effects of TZDs. It has therefore been suggested that the insulin sensitizing effects of TZDs may result from their ability to promote adipogenesis in subcutaneous adipose depots (Adams et al., ibid). Further, dominant negative mutations in the PPAR γ gene have been identified in two non-obese subjects with severe insulin resistance,
20 hypertension, and overt non-insulin dependent diabetes mellitus (NIDDM) (Barroso et al. (1998) Nature 402:880-883).

NIDDM is the most common form of diabetes mellitus, a chronic metabolic disease that affects 143 million people worldwide. NIDDM is characterized by abnormal glucose and lipid metabolism that result from a combination of peripheral insulin resistance and defective insulin
25 secretion. NIDDM has a complex, progressive etiology and a high degree of heritability. Numerous complications of diabetes including heart disease, stroke, renal failure, retinopathy, and peripheral neuropathy contribute to the high rate of morbidity and mortality.

At the molecular level, PPAR γ functions as a ligand activated transcription factor. In the presence of ligand, PPAR γ forms a heterodimer with the retinoid X receptor (RXR) which then
30 activates transcription of target genes containing one or more copies of a PPAR γ response element (PPRE). Many genes important in lipid storage and metabolism contain PPREs and have been identified as PPAR γ targets, including PEPCK, aP2, LPL, ACS, and FAT-P (Auwerx, J. (1999) Diabetologia 42:1033-1049). Multiple ligands for PPAR γ have been identified. These include a variety of fatty acid metabolites; synthetic drugs belonging to the TZD class, such as Pioglitazone and

Rosiglitazone (BRL49653); and certain non-glitazone tyrosine analogs such as GI262570 and GW1929. The prostaglandin derivative 15-dPGJ2 is a potent endogenous ligand for PPAR γ .

Expression of PPAR γ is very high in adipose but barely detectable in skeletal muscle, the primary site for insulin stimulated glucose disposal in the body. PPAR γ is also moderately expressed in large intestine, kidney, liver, vascular smooth muscle, hematopoietic cells, and macrophages. The high expression of PPAR γ in adipose suggests that the insulin sensitizing effects of TZDs may result from alterations in the expression of one or more PPAR γ regulated genes in adipose tissue.

Identification of PPAR γ target genes will contribute to better drug design and the development of novel therapeutic strategies for diabetes, obesity, and other conditions.

Systematic attempts to identify PPAR γ target genes have been made in several rodent models of obesity and diabetes (Suzuki et al. (2000) Jpn. J. Pharmacol. 84:113-123; Way et al. (2001) Endocrinology 142:1269-1277). However, a serious drawback of the rodent gene expression studies is that significant differences exist between human and rodent models of adipogenesis, diabetes, and obesity (Taylor (1999) Cell 97:9-12; Gregoire et al. (1998) Physiol. Reviews 78:783-809). Therefore, an unbiased approach to identifying TZD regulated genes in primary cultures of human tissues is necessary to fully elucidate the molecular basis for diseases associated with PPAR γ activity.

The majority of research in adipocyte biology to date has been done using transformed mouse preadipocyte cell lines. The culture condition, which stimulates mouse preadipocyte differentiation is different from that for inducing human primary preadipocyte differentiation. In addition, primary cells are diploid and may therefore reflect the *in vivo* context better than aneuploid cell lines. Understanding the gene expression profile during adipogenesis in human will lead to understanding the fundamental mechanism of adiposity regulation. Furthermore, through comparing the gene expression profiles of adipogenesis between donor with normal weight and donor with obesity, identification of crucial genes, potential drug targets for obesity and type II diabetes, will be possible.

The present invention provides for a composition comprising a plurality of cDNAs for use in detecting changes in expression of genes encoding proteins that are associated with cardiovascular tissue. Such a composition can be employed for the diagnosis, prognosis or treatment of a cardiovascular disorder and other disorders, such as atherosclerosis, hypertension, and complications of thrombolysis, balloon angioplasty, and coronary artery bypass graft surgery, correlated with differential gene expression. The present invention satisfies a need in the art in that it provides a set of differentially expressed genes which may be used entirely or in part to diagnose, to stage, to treat, or to monitor the progression or treatment of a subject with a disorder such as atherosclerosis.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of immune system disorders, neurological disorders,

developmental disorders, connective tissue disorders, and cell proliferative disorders, including cancer.

SUMMARY OF THE INVENTION

5 Various embodiments of the invention provide purified polypeptides, cell adhesion and extracellular matrix proteins, referred to collectively as 'CADECM' and individually as 'CADECM-1,' 'CADECM-2,' 'CADECM-3,' 'CADECM-4,' 'CADECM-5,' 'CADECM-6,' 'CADECM-7,' 'CADECM-8,' 'CADECM-9,' 'CADECM-10,' 'CADECM-11,' 'CADECM-12,' 'CADECM-13,' 'CADECM-14,' 'CADECM-15,' 'CADECM-16,' 'CADECM-17,' 'CADECM-18,' 'CADECM-19,'
 10 'CADECM-20,' 'CADECM-21,' 'CADECM-22,' 'CADECM-23,' 'CADECM-24,' 'CADECM-25,' 'CADECM-26,' 'CADECM-27,' 'CADECM-28,' 'CADECM-29,' 'CADECM-30,' 'CADECM-31,' 'CADECM-32,' 'CADECM-33,' 'CADECM-34,' 'CADECM-35,' 'CADECM-36,' and 'CADECM-37' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for
 15 utilizing the purified cell adhesion and extracellular matrix proteins and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified cell adhesion and extracellular matrix proteins and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

20 An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected
 25 from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-37.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected
 30 from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an

immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-37. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:38-74.

5 Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group
10 consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

15 Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group
20 consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering
25 the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an
30 amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37.

Still yet another embodiment provides an isolated polynucleotide selected from the group

consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid

sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional CADECM, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional CADECM, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment

provides a method of treating a disease or condition associated with overexpression of functional CADECM, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

“CADECM” refers to the amino acid sequences of substantially purified CADECM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of CADECM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CADECM either by directly

interacting with CADECM or by acting on components of the biological pathway in which CADECM participates.

An “allelic variant” is an alternative form of the gene encoding CADECM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding CADECM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CADECM or a polypeptide with at least one functional characteristic of CADECM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CADECM, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding CADECM. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CADECM. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CADECM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms “amino acid” and “amino acid sequence” can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other

nucleic acid amplification technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of CADECM. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CADECM either by directly interacting with CADECM or by acting on components of the biological pathway in which CADECM participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CADECM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term “antigenic determinant” refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term “aptamer” refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term “intramer” refers to an aptamer which is expressed *in vivo*. For example, a vaccinia

virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term “spiegelmer” refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed
5 nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term “antisense” refers to any composition capable of base-pairing with the “sense” (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone
10 linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a
15 naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation “negative” or “minus” can refer to the antisense strand, and the designation “positive” or “plus” can refer to the sense strand of a reference DNA molecule.

The term “biologically active” refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, “immunologically active” or “immunogenic”
20 refers to the capability of the natural, recombinant, or synthetic CADECM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

“Complementary” describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement,
25 3'-TCA-5'.

A “composition comprising a given polynucleotide” and a “composition comprising a given polypeptide” can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding CADECM or fragments of CADECM may be employed as hybridization
30 probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

“Consensus sequence” refers to a nucleic acid sequence which has been subjected to repeated

DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

| | Original Residue | Conservative Substitution |
|----|------------------|---------------------------|
| | Ala | Gly, Ser |
| | Arg | His, Lys |
| 15 | Asn | Asp, Gln, His |
| | Asp | Asn, Glu |
| | Cys | Ala, Ser |
| | Gln | Asn, Glu, His |
| | Glu | Asp, Gln, His |
| 20 | Gly | Ala |
| | His | Asn, Arg, Gln, Glu |
| | Ile | Leu, Val |
| | Leu | Ile, Val |
| | Lys | Arg, Gln, Glu |
| 25 | Met | Leu, Ile |
| | Phe | His, Met, Leu, Trp, Tyr |
| | Ser | Cys, Thr |
| | Thr | Ser, Val |
| | Trp | Phe, Tyr |
| 30 | Tyr | His, Phe, Trp |
| | Val | Ile, Leu, Thr |

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide.

Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which

retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

5 A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

“Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

10 “Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A “fragment” is a unique portion of CADECM or a polynucleotide encoding CADECM which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500
20 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures,
25 may be encompassed by the present embodiments.

A fragment of SEQ ID NO:38-74 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:38-74, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:38-74 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and
30 amplification technologies and in analogous methods that distinguish SEQ ID NO:38-74 from related polynucleotides. The precise length of a fragment of SEQ ID NO:38-74 and the region of SEQ ID NO:38-74 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-37 is encoded by a fragment of SEQ ID NO:38-74. A fragment

of SEQ ID NO:1-37 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-37. For example, a fragment of SEQ ID NO:1-37 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-37. The precise length of a fragment of SEQ ID NO:1-37 and the region of SEQ ID NO:1-37 to which the fragment
5 corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A “full length” polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

10 “Homology” refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of identical residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible
15 way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the
20 MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and
25 “diagonals saved”=4. The “weighted” residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at
30 <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including “blastn,” that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called “BLAST 2 Sequences” that is used for direct pairwise comparison of two nucleotide sequences. “BLAST 2 Sequences” can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>.

The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

5 *Matrix: BLOSUM62*
 Reward for match: 1
 Penalty for mismatch: -2
 Open Gap: 5 and Extension Gap: 2 penalties
 Gap x drop-off: 50
 10 *Expect: 10*
 Word Size: 11
 Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example,
 15 over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

20 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to
 25 the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The
 30 phrases “percent similarity” and “% similarity,” as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap
 5 penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for
 10 example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

15 *Word Size: 3*

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for
 20 instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain
 25 DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

30 "Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the

stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%.

Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

“Immune response” can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

5 An “immunogenic fragment” is a polypeptide or oligopeptide fragment of CADECM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term “immunogenic fragment” also includes any polypeptide or oligopeptide fragment of CADECM which is useful in any of the antibody production methods disclosed herein or known in the art.

10 The term “microarray” refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

 The terms “element” and “array element” refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

 The term “modulate” refers to a change in the activity of CADECM. For example,
15 modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CADECM.

 The phrases “nucleic acid” and “nucleic acid sequence” refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the
20 antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

 “Operably linked” refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where
25 necessary to join two protein coding regions, in the same reading frame.

 “Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript
30 elongation, and may be pegylated to extend their lifespan in the cell.

 “Post-translational modification” of an CADECM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of CADECM.

“Probe” refers to nucleic acids encoding CADECM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may

also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing CADECM, nucleic acids encoding CADECM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

5 The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide
10 comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

 The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other
15 components with which they are naturally associated.

 A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

 "Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,
20 microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

 A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

 "Transformation" describes a process by which exogenous DNA is introduced into a recipient
25 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term
30 "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic

acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) *Science* 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for

example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

5 THE INVENTION

Various embodiments of the invention include new human cell adhesion and extracellular matrix proteins (CADECM), the polynucleotides encoding CADECM, and the use of these compositions for the diagnosis, treatment, or prevention of immune system disorders, neurological disorders, developmental disorders, connective tissue disorders, and cell proliferative disorders,
10 including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an
15 Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as
20 identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers
25 (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1
30 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Accelrys, Burlington MA).

Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are cell adhesion and extracellular matrix proteins. For example, SEQ ID NO:4 is 97% identical, from residue A81 to residue A1102, to human latent transforming growth factor-beta binding protein 4S (GenBank ID g3327808) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:4 also has homology to proteins that are localized to the extracellular matrix, that function as small molecule-binding proteins, and are latent transforming growth factor-beta binding proteins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:4 also contains 20 EGF-like domains and three TB domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:4 is a latent transforming growth factor-beta binding protein.

In an alternative example, SEQ ID NO:10 is 100% identical, from residue M1 to residue T430, to protocadherin beta 2 (GenBank ID g5457039) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.3e-253, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:10 also has homology to proteins that are localized to neural proteins, have adhesion function, and are cadherins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:10 also contains a cadherin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:10 is a cadherin protein.

In an alternative example, For example, SEQ ID NO:15 is 100% identical, from residue M1 to residue Q1012, to human protocadherin-9 (GenBank ID g9845485) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:15 also has homology to proteins that are localized to cell adhesion molecules, involved in cell-cell recognition in the central nervous system, and are cadherins, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:15 also contains a cadherin domain as

determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN, and other BLAST analyses provide further corroborative evidence that SEQ ID NO:15 is a cadherin.

5 In an alternative example, SEQ ID NO:17 is 99.9% identical, from residue M1 to residue T1805, to human integrin beta 4 subunit (GenBank ID g33957) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:17 also has homology to proteins that are localized to the plasma membrane, function as a
10 receptor, and are integrin beta 4 subunits, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:17 also contains a fibronectin type III domain and an integrin beta chain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, PROFILESCAN, and other BLAST analyses provide further corroborative evidence that
15 SEQ ID NO:17 is an integrin beta 4 subunit.

In an alternative example, SEQ ID NO:22 is 99% identical, from residue L38 to residue C715, to human integrin beta-subunit (GenBank ID g9446402) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ
20 ID NO:22 also has homology to proteins that are localized to the plasma membrane, mediate cell interactions with the extracellular matrix, are involved in platelet aggregation, cell adhesion, blood coagulation, and cell proliferation and differentiation, and are integrin beta subunits, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:22 also contains an integrin beta subunit domain as determined by searching for statistically significant matches in the hidden Markov
25 model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Analysis by TMHMMER indicates the presence of a transmembrane domain from I635 to W657. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses and BLAST analyses of the PRODOM and DOMO data bases provide further corroborative evidence that SEQ ID NO:22 is an integrin subunit.

30 In an alternative example, SEQ ID NO:29 is 100% identical, from residue M1 to residue I834, to human MUCDHL-FL (GenBank ID g10334774) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:29 also has homology to proteins that are localized to the basolateral plasma membrane and are similar to cell

adhesion molecules such as cadherin and mucin, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:29 also contains a cadherin domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS, MOTIFS, and BLAST analyses provide further corroborative evidence that SEQ ID NO:29 is a cadherin.

In an alternative example, SEQ ID NO:35 is 99% identical, from residue M1 to residue A859, to human N-CAM (GenBank ID g632776) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:35 also has homology to proteins that are members of the immunoglobulin superfamily, are localized to the plasma membrane, are important for cell adhesion between neurons and at neuromuscular junctions, and function to regulate axonal outgrowth and neuronal remodeling, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:35 also contains several immunoglobulin domains and two fibronectin type III domains, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM and SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and additional BLAST analyses using the PRODOM and DOMO databases provide further corroborative evidence that SEQ ID NO:35 is a neural cell adhesion molecule.

SEQ ID NO:1-3, SEQ ID NO:5-9, SEQ ID NO:11-14, SEQ ID NO:16, SEQ ID NO:18-21, SEQ ID NO:23-28, SEQ ID NO:30-34, and SEQ ID NO:36-37 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-37 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:38-74 or that distinguish between SEQ ID NO:38-74 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA

libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation “ENST”). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a “stitched” sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAA_gBBBBB_1_N is a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

| Prefix | Type of analysis and/or examples of programs |
|----------------|--|
| GNN, GFG, ENST | Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK). |
| GBI | Hand-edited analysis of genomic sequences. |
| FL | Stitched or stretched genomic sequences (see Example V). |

| | |
|------|---|
| INCY | Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript. |
|------|---|

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses CADECM variants. Various embodiments of CADECM variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the CADECM amino acid sequence, and can contain at least one functional or structural characteristic of CADECM.

Various embodiments also encompass polynucleotides which encode CADECM. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:38-74, which encodes CADECM. The polynucleotide sequences of SEQ ID NO:38-74, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with

uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding CADECM. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding CADECM. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:38-74 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:38-74. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of CADECM.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding CADECM. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding CADECM, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing of exons during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding CADECM over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding CADECM. For example, a polynucleotide comprising a sequence of SEQ ID NO:58, and a polynucleotide comprising a sequence of SEQ ID NO:59 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:60, and a polynucleotide comprising a sequence of SEQ ID NO:64 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:62, and a polynucleotide comprising a sequence of SEQ ID NO:63 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:65, and a polynucleotide comprising a sequence of SEQ ID NO:74 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:51, a polynucleotide comprising a sequence of SEQ ID NO:66, a polynucleotide comprising a sequence of SEQ ID NO:67, and a polynucleotide comprising a sequence of SEQ ID NO:68 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of CADECM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CADECM, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be

produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CADECM, and all such variations are to be
5 considered as being specifically disclosed.

Although polynucleotides which encode CADECM and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring CADECM under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding CADECM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally
10 occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CADECM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater
15 half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode CADECM and CADECM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to
20 introduce mutations into a polynucleotide encoding CADECM or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:38-74 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.*
25 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied
30 Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems).

Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and
5 Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding CADECM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic
10 DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known
15 sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may
20 use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of
25 about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence
30 into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the

emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments
5 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode CADECM may be cloned in recombinant DNA molecules that direct expression of CADECM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally
10 equivalent polypeptides may be produced and used to express CADECM.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter CADECM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be
15 used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No.
20 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CADECM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then
25 subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are
30 optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding CADECM may be synthesized, in whole or

in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, CADECM itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of CADECM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active CADECM, the polynucleotides encoding CADECM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding CADECM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding CADECM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding CADECM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding CADECM and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8;

Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding CADECM. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding CADECM. For example, routine cloning, subcloning, and propagation of polynucleotides encoding CADECM can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding CADECM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of CADECM are needed, e.g. for the production of antibodies, vectors which direct high level expression of CADECM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CADECM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such

vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184).

5 Plant systems may also be used for expression of CADECM. Transcription of polynucleotides encoding CADECM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; 10 Broglie, R. et al. (1984) *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases 15 where an adenovirus is used as an expression vector, polynucleotides encoding CADECM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CADECM in host cells (Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659). In addition, transcription enhancers, such as the Rous 20 sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino 25 polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of CADECM in cell lines is preferred. For example, polynucleotides encoding CADECM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate 30 vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CADECM is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding CADECM can be identified by the absence of marker gene function.

Alternatively, a marker gene can be placed in tandem with a sequence encoding CADECM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding CADECM and that express CADECM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CADECM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CADECM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect.

IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CADECM include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding CADECM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding CADECM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CADECM may be designed to contain signal sequences which direct secretion of CADECM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding CADECM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CADECM protein containing a heterologous moiety that can be recognized by a commercially available antibody may

facilitate the screening of peptide libraries for inhibitors of CADECM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CADECM encoding sequence and the heterologous protein sequence, so that CADECM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled CADECM may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

CADECM, fragments of CADECM, or variants of CADECM may be used to screen for compounds that specifically bind to CADECM. One or more test compounds may be screened for specific binding to CADECM. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to CADECM. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of CADECM can be used to screen for binding of test compounds, such as antibodies, to CADECM, a variant of CADECM, or a combination of CADECM and/or one or more variants CADECM. In an embodiment, a variant of CADECM can be used to screen for compounds that bind to a variant of CADECM, but not to CADECM having the exact sequence of a sequence of SEQ ID NO:1-37. CADECM variants used to perform such screening can have a range of about 50% to about 99% sequence identity to CADECM, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to CADECM can be closely related to the natural ligand of CADECM, e.g., a ligand or fragment thereof, a natural

substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor CADECM (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

5 In other embodiments, a compound identified in a screen for specific binding to CADECM can be closely related to the natural receptor to which CADECM binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for CADECM which is capable of propagating a signal, or a decoy receptor for CADECM which is not capable of propagating a signal
10 (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion
15 of human IgG₁ (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to CADECM, fragments of CADECM, or variants of CADECM. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of CADECM. In one embodiment, an antibody can be
20 selected such that its binding specificity allows for preferential identification of specific fragments or variants of CADECM. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of CADECM.

In an embodiment, anticalins can be screened for specific binding to CADECM, fragments of
25 CADECM, or variants of CADECM. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-
30 engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit

CADECM involves producing appropriate cells which express CADECM, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing CADECM or cell membrane fractions which contain CADECM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either CADECM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with CADECM, either in solution or affixed to a solid support, and detecting the binding of CADECM to the compound.

Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

CADECM, fragments of CADECM, or variants of CADECM may be used to screen for compounds that modulate the activity of CADECM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for CADECM activity, wherein CADECM is combined with at least one test compound, and the activity of CADECM in the presence of a test compound is compared with the activity of CADECM in the absence of the test compound. A change in the activity of CADECM in the presence of the test compound is indicative of a compound that modulates the activity of CADECM. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising CADECM under conditions suitable for CADECM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of CADECM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a

plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding CADECM or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding CADECM may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding CADECM can also be used to create “knockin” humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding CADECM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress CADECM, e.g., by secreting CADECM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CADECM and cell adhesion and extracellular matrix proteins. In addition,

examples of tissues expressing CADECM can be found in Table 6 and can also be found in Example XI. Therefore, CADECM appears to play a role in immune system disorders, neurological disorders, developmental disorders, connective tissue disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased CADECM expression or activity, it is desirable to decrease the expression or activity of CADECM. In the treatment of disorders associated with decreased CADECM expression or activity, it is desirable to increase the expression or activity of CADECM.

Therefore, in one embodiment, CADECM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADECM. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases

including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down

- 5 syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia,
- 10 diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental
- 15 retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a connective tissue disorder, such as osteogenesis imperfecta, Ehlers-
- 20 Danlos syndrome, chondrodysplasias, Marfan syndrome, Alport syndrome, familial aortic aneurysm, achondroplasia, mucopolysaccharidoses, osteoporosis, osteopetrosis, Paget's disease, rickets, osteomalacia, hyperparathyroidism, renal osteodystrophy, osteonecrosis, osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondroma, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defect, nonossifying fibroma, fibrous
- 25 dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing's sarcoma, primitive neuroectodermal tumor, giant cell tumor, osteoarthritis, rheumatoid arthritis, ankylosing spondyloarthritis, Reiter's syndrome, psoriatic arthritis, enteropathic arthritis, infectious arthritis, gout, gouty arthritis, calcium pyrophosphate crystal deposition disease, ganglion, synovial cyst, villonodular synovitis, systemic sclerosis, Dupuytren's contracture, hepatic fibrosis, lupus
- 30 erythematosus, mixed connective tissue disease, epidermolysis bullosa simplex, bullous congenital ichthyosiform erythroderma (epidermolytic hyperkeratosis), non-epidermolytic and epidermolytic palmoplantar keratoderma, ichthyosis bullosa of Siemens, pachyonychia congenita, and white sponge nevus; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal

nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing CADECM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADECM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified CADECM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADECM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CADECM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CADECM including, but not limited to, those listed above.

In a further embodiment, an antagonist of CADECM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CADECM. Examples of such disorders include, but are not limited to, those immune system disorders, neurological disorders, developmental disorders, connective tissue disorders, and cell proliferative disorders, including cancer described above. In one aspect, an antibody which specifically binds CADECM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CADECM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CADECM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CADECM including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CADECM may be produced using methods which are generally known in the art. In particular, purified CADECM may be used to produce antibodies or to screen libraries of

pharmaceutical agents to identify those which specifically bind CADECM. Antibodies to CADECM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e.,
5 those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels,
10 dromedaries, llamas, humans, and others may be immunized by injection with CADECM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol.
15 Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CADECM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or
20 fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of CADECM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CADECM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not
25 limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the
30 splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CADECM-specific single

chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte
5 population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for CADECM may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced
10 by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired
15 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CADECM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CADECM epitopes is generally used, but a competitive binding assay
20 may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CADECM. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of CADECM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.
25 The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CADECM epitopes, represents the average affinity, or avidity, of the antibodies for CADECM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CADECM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in
30 immunoassays in which the CADECM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CADECM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to

Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CADECM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding CADECM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding CADECM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CADECM (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) *Blood* 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) *Br. Med. Bull.* 51:217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87:1308-1315; Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding CADECM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene*

Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in CADECM expression or regulation causes disease, the expression of CADECM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in CADECM are treated by constructing mammalian expression vectors encoding CADECM and introducing these vectors by mechanical means into CADECM-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of CADECM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). CADECM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding CADECM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to CADECM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding CADECM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding CADECM to cells which have one or more genetic abnormalities with respect to the expression of CADECM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene

therapy”), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; *Annu. Rev. Nutr.* 19:511-544) and Verma, I.M. and N. Somia (1997; *Nature* 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding CADECM to target cells which have one or more genetic abnormalities with respect to the expression of CADECM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing CADECM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca (“Herpes simplex virus strains for gene transfer”), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; *J. Virol.* 73:519-532) and Xu, H. et al. (1994; *Dev. Biol.* 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding CADECM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for CADECM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of CADECM-coding RNAs and the synthesis of high levels of CADECM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of

alphaviruses will allow the introduction of CADECM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding CADECM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding CADECM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible

modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. SiRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. SiRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

SiRNA can either be generated indirectly by introduction of dsRNA into the targeted cell, or directly by mammalian transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable SiRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The

selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected SiRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene, can be determined by northern analysis methods using, for example, the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding CADECM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased CADECM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding CADECM may be therapeutically useful, and in the treatment of disorders associated with decreased CADECM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding CADECM may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding CADECM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding CADECM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding CADECM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of
5 Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of CADECM, antibodies to CADECM, and mimetics, agonists, antagonists, or inhibitors of CADECM.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral,
10 intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-
15 acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

20 Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising CADECM or fragments thereof. For example, liposome preparations
25 containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, CADECM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

30 For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CADECM or fragments thereof, antibodies of CADECM, and agonists, antagonists or inhibitors of CADECM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CADECM may be used for the diagnosis of disorders characterized by expression of CADECM, or in assays to monitor patients being treated with CADECM or agonists, antagonists, or inhibitors of CADECM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CADECM include methods which utilize the antibody and a label to detect CADECM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in

the art and may be used.

A variety of protocols for measuring CADECM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CADECM expression. Normal or standard values for CADECM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to CADECM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CADECM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding CADECM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CADECM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CADECM, and to monitor regulation of CADECM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding CADECM or closely related molecules may be used to identify nucleic acid sequences which encode CADECM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CADECM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CADECM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:38-74 or from genomic sequences including promoters, enhancers, and introns of the CADECM gene.

Means for producing specific hybridization probes for polynucleotides encoding CADECM include the cloning of polynucleotides encoding CADECM or CADECM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ³²P or ³⁵S, or by enzymatic labels,

such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding CADECM may be used for the diagnosis of disorders associated with expression of CADECM. Examples of such disorders include, but are not limited to, an immune system disorder, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVID), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder, such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and

toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial

5 frontotemporal dementia; a developmental disorder, such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and

10 neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a connective tissue disorder, such as osteogenesis imperfecta, Ehlers-Danlos syndrome, chondrodysplasias, Marfan syndrome, Alport syndrome, familial aortic aneurysm, achondroplasia, mucopolysaccharidoses, osteoporosis, osteopetrosis, Paget's disease, rickets,

15 osteomalacia, hyperparathyroidism, renal osteodystrophy, osteonecrosis, osteomyelitis, osteoma, osteoid osteoma, osteoblastoma, osteosarcoma, osteochondroma, chondroma, chondroblastoma, chondromyxoid fibroma, chondrosarcoma, fibrous cortical defect, nonossifying fibroma, fibrous dysplasia, fibrosarcoma, malignant fibrous histiocytoma, Ewing's sarcoma, primitive neuroectodermal tumor, giant cell tumor, osteoarthritis, rheumatoid arthritis, ankylosing

20 spondyloarthritis, Reiter's syndrome, psoriatic arthritis, enteropathic arthritis, infectious arthritis, gout, gouty arthritis, calcium pyrophosphate crystal deposition disease, ganglion, synovial cyst, villonodular synovitis, systemic sclerosis, Dupuytren's contracture, hepatic fibrosis, lupus erythematosus, mixed connective tissue disease, epidermolysis bullosa simplex, bullous congenital ichthyosiform erythroderma (epidermolytic hyperkeratosis), non-epidermolytic and epidermolytic

25 palmoplantar keratoderma, ichthyosis bullosa of Siemens, pachyonychia congenita, and white sponge nevus; and a cell proliferative disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma,

30 and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

Polynucleotides encoding CADECM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like

assays; and in microarrays utilizing fluids or tissues from patients to detect altered CADECM expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding CADECM may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding CADECM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding CADECM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CADECM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding CADECM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CADECM may involve the use of PCR. These oligomers may be chemically synthesized, generated

enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding CADECM, or a fragment of a polynucleotide complementary to the polynucleotide encoding CADECM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or
5 quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding CADECM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism
10 (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding CADECM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel
15 electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-
20 based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16
25 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that
30 influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating

genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) *Trends Mol. Med.* 7:507-512; Kwok, P.-Y. and Z. Gu (1999) *Mol. Med. Today* 5:538-543; Nowotny, P. et al. (2001) *Curr. Opin. Neurobiol.* 11:637-641).

Methods which may also be used to quantify the expression of CADECM include

5 radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid
10 quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic
15 variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment
20 regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, CADECM, fragments of CADECM, or antibodies specific for CADECM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as
25 described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at
30 a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present

invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*,
5 as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed
10 molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number
15 of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids
20 in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

25 In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with
30 levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected

individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using
5 two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot
10 is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed
15 by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for CADECM to quantify the levels of CADECM expression. In one embodiment, the antibodies are used as elements
20 on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of
25 fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be
30 useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological

sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding CADECM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and

D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM)

5 World Wide Web site. Correlation between the location of the gene encoding CADECM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps.

10 Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23,
15 any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CADECM, its catalytic or immunogenic fragments,
20 or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CADECM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds
25 having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CADECM, or fragments thereof, and washed. Bound CADECM is then detected by methods well known in the art. Purified CADECM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,
30 non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CADECM specifically compete with a test compound for binding CADECM. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CADECM.

In additional embodiments, the nucleotide sequences which encode CADECM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/334,343, U.S. Ser. No. 60/340,278, U.S. Ser. No. 60/345,069, U.S. Ser. No. 60/351,352, U.S. Ser. No. 60/369,128, U.S. Ser. No. 60/357,168, and U.S. Ser. No. 60/370,802, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or

enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI

protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

5 The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and
10 BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as
15 SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide
20 sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive
25 the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM;
30 and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between

aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:38-74. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative cell adhesion and extracellular matrix proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94; Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode cell adhesion and extracellular matrix proteins, the encoded polypeptides were analyzed by querying against PFAM models for cell adhesion and extracellular matrix proteins. Potential cell adhesion and extracellular matrix proteins were also identified by homology to Incyte cDNA sequences that had been annotated as cell adhesion and extracellular matrix proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or

public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

5 "Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm
10 based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic
15 sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or
20 genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

25 "Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST
30 analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for

homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of CADECM Encoding Polynucleotides

5 The sequences which were used to assemble SEQ ID NO:38-74 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:38-74 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available
10 from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

 Map locations are represented by ranges, or intervals, of human chromosomes. The map
15 position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation
20 hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

 Heritable forms of lung carcinoma have not been reported and thus, identification of relevant
25 disease-associated genes through conventional linkage analysis is not possible. However, several studies of sporadic nonsmall cell lung carcinoma (NSCLC) tumors have reported loss of heterozygosity (LOH) in regions of chromosome 11 suggesting the presence of one or more tumor suppressor genes (Bepler, G. and Garcia-Blanco, M.A. (1994) Proc. Natl. Acad. Sci. U.S.A. 91:5513-7; Iizuka, M. (1995) Genes, Chromosomes & Cancer 13:40-46; Rasio, D. (1995) Cancer Research
30 55:3988-91). In a study of 79 patients with lung cancer, Iizuka and coworkers found that 11q14-11q24.2 was deleted in many of the lung tumors studied. Mapping of this region with additional markers showed that the region of chromosome 11q bounded by markers D11S939 and D11S938 was commonly deleted (Iizuka, et al., *supra*). In another study it was shown that human A549 NSCLC cells transformed with a human-derived YAC clone containing a region of chromosome 11q within

the region bounded by D11S939 and D11S938, exhibited little or no increase in cell number (versus control cells whose number increased 5-10-fold in the same 5 day period). Further studies of these hybrid cells showed a decrease in tumorigenicity and an increase in latency following injection into athymic, nude mice, as compared with mice injected with control A549 cells. These studies suggest the presence of a tumor suppressor gene within this region of chromosome 11q and support the association of LOH in this region with NSCLC.

Restriction fragment length polymorphism (RFLP) markers shown to be near regions of DNA known as sequence-tagged sites (STS), have been mapped to NT_Contigs generated by the Human Genome Project using ePCR (Schuler, G.D. (1997) *Genome Research* 7: 541-550, and (1998) *Trends Biotechnol.* 16(11):456-459). Contigs containing regions of DNA with known disease-associated markers are therefore used to identify CADECM sequences that map to disease-associated regions of the genome.

Polynucleotides encoding CADECM were mapped to NT_Contigs. Contigs longer than 1Mb were broken into subcontigs of 1Mb length with overlapping sections of 100kb. A preliminary step used an algorithm, similar to MEGABLAST, to define the mRNA sequence /masked genomic DNA contig pairings. The cDNA/genomic pairings identified by the first algorithm were confirmed, and the CADECM polynucleotides mapped to DNA contigs, using SIM4 (Florea, L. et al. (1998) *Genome Res.* 8:967-74, version May 2000) which had been optimized for high throughput processing and strand assignment confidence). The SIM4 output of the mRNA sequence/genomic contig pairs was further processed to determine the correct location of the CADECM polynucleotides on the genomic contig, as well as their strand identity.

SEQ ID NO:72 was mapped to NT_Contig GBI:NT_009151_018.8 from Genbank release February 2002, covering a 5.5 Mb region of the genome that also contains lung cancer-associated genetic markers D11S939 and D11S938. The maximum distance between SEQ ID NO:72 and markers D11S939 and D11S938, therefore, is 5.5 Mb. Thus, SEQ ID NO:72 is in proximity with genetic markers shown to consistently associate with lung cancer. Therefore, in various embodiments, SEQ ID NO:72 can be used for one or more of the following: i) determination of LOH in persons with lung cancer in the lung cancer disease region at 11q12-24.2, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar.

- 5 The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

- 10 The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair
 15 (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced
 20 either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

- Alternatively, polynucleotides encoding CADECM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at
 25 least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas;
 30 respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided
 35 by the total number of libraries across all categories. The resulting percentages reflect the tissue- and

disease-specific expression of cDNA encoding CADECM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of CADECM Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun

sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and
5 transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step
10 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers
15 and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

20 IX. Identification of Single Nucleotide Polymorphisms in CADECM Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:38-74 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the
25 identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original
30 chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in

immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:38-74 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ - 32 P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed.

(1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) *Science* 270:467-470; Shalon, D. et al. (1996) *Genome Res.* 6:639-645; Marshall, A. and J. Hodgson (1998) *Nat. Biotechnol.* 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is

focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

5 In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is
10 typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that
15 location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

20 The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and
25 measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The
30 software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

For example, SEQ ID NO:39 and SEQ ID NO:41-42 showed differential expression in tumorous tissue versus non-tumorous tissues and in various treated versus non-treated cell lines, as determined by microarray analysis. The expression of cDNAs from lung tumor tissue from several donors was compared with that of normal lung tissue from the same donor, respectively. Array elements that exhibited about at least a two-fold change in expression and a signal intensity over 250 units, a signal-to-background ratio of at least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

The expression of SEQ ID NO:39 was decreased at least two-fold in lung adenocarcinoma and increased at least two-fold in lung squamous cell carcinoma when matched with normal tissue from the same donor. In one case, the tumorous lung adenocarcinoma tissue was obtained from lung of a 71-year old female. In the other case, the tumorous lung squamous cell carcinoma tissue was obtained from lung of a 68-year old female. Normal tissue was obtained from grossly uninvolved lung tissue from the same donor, respectively. Therefore, SEQ ID NO:39 is useful in diagnostic assays for lung adenocarcinoma. Matched normal and tumorigenic colon tissue samples are provided by the Roy Castle International Centre for Lung Cancer Research (Liverpool, UK).

Further, SEQ ID NO:39 showed differential expression in senescent (passage 8) and pre-senescent (passage 7) versus non-senescent progenitor PrEC cells (passage 3). PrEC are primary prostate epithelial cells isolated from a normal donor and were grown in the optimal growth media to 70-80% confluence prior to harvesting. Therefore, SEQ ID NO:39 is useful as a diagnostic marker or as a potential therapeutic target for cancer.

Yet further, SEQ ID NO:39 showed at least two-fold decreased expression in C3A cells treated with a variety of steroids including beclomethasone, prednisone, and dexamethasone, versus untreated C3A cells, as determined by microarray analysis. The effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug. For example, the human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α -fetoprotein; iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an *in vitro* model of the mature human liver (Mickelson et al. (1995) Hepatology 22:866-875; Nagendra et al. (1997) Am. J. Physiol. 272:G408-G416). Therefore, SEQ ID NO:39 is useful in the diagnosis of and as a therapeutic target for inflammatory diseases and humoral immune

response.

The expression of SEQ ID NO:41 was decreased at least 3.1-fold in lung adenocarcinoma when matched with normal tissue from the same donor. The tumorous lung tissue was obtained from lung of a 60-year old donor with moderately differentiated adenocarcinoma of the right lung. Normal
5 tissue was obtained from grossly uninvolved right lung tissue from the same donor.

The expression of SEQ ID NO:41 also was decreased at least 2.7-fold in breast carcinoma when matched with normal tissue from the same donor. The tumorous breast tissue was obtained from the right breast of a 43-year old female with invasive lobular carcinoma in situ which had metastasized to two out of 13 lymph nodes. Normal tissue was obtained from grossly uninvolved
10 breast tissue from the same donor.

The expression of SEQ ID NO:41 also was decreased at least two-fold in colon adenocarcinoma when matched with normal tissue from the same donor. The tumorous colon tissue was obtained from the colon of a 67-year old male with moderately differentiated adenocarcinoma. Normal tissue was obtained from grossly uninvolved colon tissue from the same donor.

The expression of SEQ ID NO:41 also was decreased at least two-fold in ovarian adenocarcinoma when matched with normal tissue from the same donor. The tumorous tissue was obtained from ovary of a 79-year old female with ovarian adenocarcinoma. Normal tissue was obtained from grossly uninvolved ovarian tissue from the same donor. Therefore, SEQ ID NO:41 is useful in diagnostic assays for adenocarcinoma. Matched normal and tumorigenic lung, breast, colon,
15 and ovary tissue samples are provided by the Huntsman Cancer Institute (Salt Lake City, UT).
20

Further, the expression of SEQ ID NO:42 was increased at least 2.8-fold in lung adenocarcinoma when matched with normal tissue from the same donor. The tumorous lung tissue was obtained from lung of a 66-year old female with lung adenocarcinoma. Normal tissue was obtained from grossly uninvolved tissue from the same donor. The expression of SEQ ID NO:42 also
25 was increased at least 2.3-fold in lung squamous cell carcinoma in five donors when matched with normal tissue from the same donor. The tumorous lung tissue was obtained from the lung of a 75-year-old female, two different 73-year-old males, a 68-year-old female, and a 66-year-old male, all with lung squamous cell carcinoma. Normal tissue was obtained from grossly uninvolved lung tissue from the same donor. Therefore, SEQ ID NO:42 is useful in diagnostic assays for lung
30 adenocarcinoma and squamous cell carcinoma. Matched normal and tumorigenic lung samples were obtained from the Roy Castle International Centre for Lung Cancer Research (Liverpool, UK).

The expression of SEQ ID NO:46 was found to be differentially expressed in human breast tumor tissue as compared to normal breast tissue from the same donor. In an alternative example, the expression of SEQ ID NO:46 was also found to be differentially expressed in human lung tumor

tissue as compared to normal lung tissue from the same donor. SEQ ID NO:46 was underexpressed by at least two-fold in both breast and lung tumor tissue as compared to normal breast and lung tissues from the same donors, respectively. In addition, the expression of SEQ ID NO:46 was differentially expressed. These experiments indicate that SEQ ID NO:46 exhibited significant differential expression patterns using microarray techniques, and further establishes its utility as a diagnostic marker or therapeutic agent which may be useful in a variety of conditions and diseases involving cell adhesion and extracellular matrix proteins, including cancer and atherosclerosis.

The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) HMEC, a primary breast epithelial cell line isolated from a normal donor (Clonetics, San Diego CA), b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, g) MDA-MB-468, a breast adenocarcinoma cell line isolated from the pleural effusion of a 51-year-old female with metastatic adenocarcinoma of the breast, h) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, i) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, j) BT-20, a breast carcinoma cell line derived in vitro from cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, k) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year-old female, and l) MDA-mb-435S, a spindle-shaped strain that evolved from the parent line (435) isolated by R. Cailleau from pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast.

For example, the expression of SEQ ID NO:54 was decreased between two-fold and 11-fold in nine (BT-20, BT-474, BT-483, Hs 578T, MCF7, T-47D, Sk-BR-3, MDA-mb-231, and MDA-mb-435S) of ten human breast tumor cell lines (described above) when compared to a nonmalignant mammary epithelial cell line (HMEC) or a breast mammary gland cell line isolated from a woman with fibrocystic breast disease (MCF-10A). Cell lines were grown in various media ranging from media with growth factors and hormones, to basal media in the absence of growth factors and hormones. Therefore, SEQ ID NO:54 is useful as a diagnostic marker or as a potential therapeutic

target for early and late stage breast cancer.

In another example, the expression of SEQ ID NO:54 was decreased at least 2.3-fold in cancerous colon tissue compared to normal tissue from the same donor. Sigmoid colon tissue was obtained from a 48-year-old female and matched with normal sigmoid colon tissue obtained from
5 grossly uninvolved tissue from the same donor. Therefore, SEQ ID NO:54 is useful in diagnostic assays for sigmoid colon cancer.

In yet another example, the expression of SEQ ID NO:54 was increased at least 2.6-fold in cancerous lung tissue compared to normal tissue from the same donor. Lung squamous cell carcinoma tissue was obtained from a 73-year-old male and matched with normal lung tissue obtained
10 from grossly uninvolved tissue from the same donor. Therefore, SEQ ID NO:54 is useful in diagnostic assays for lung squamous cell carcinoma.

In a further example, the expression of SEQ ID NO:54 was increased at least two-fold in cancerous ovarian tissue compared to normal tissue from the same donor. Ovary adenocarcinoma tissue was obtained from a 79-year-old female and matched with normal ovary tissue obtained from
15 the same donor. Therefore, SEQ ID NO:54 is useful in diagnostic assays for ovarian adenocarcinoma.

Matched normal and tumorigenic colon and ovary tissue samples are provided by the Huntsman Cancer Institute (Salt Lake City, UT). Matched normal and tumorigenic lung tissue samples are provided by the Roy Castle International Centre for Lung Cancer Research, Liverpool UK).

20 As with most tumors, prostate cancer develops through a multistage progression ultimately resulting in an aggressive tumor phenotype. The initial step in tumor progression involves the hyperproliferation of normal luminal and/or basal epithelial cells. Androgen responsive cells become hyperplastic and evolve into early-stage tumors. Although early-stage tumors are often androgen sensitive and respond to androgen ablation, a population of androgen independent cells evolve from
25 the hyperplastic population. These cells represent a more advanced form of prostate tumor that may become invasive and potentially become metastatic to the bone, brain, or lung. PrEC is a primary prostate epithelial cell line isolated from a normal donor. It was obtained from Clinomics Corporation (Walkersville MD). PZ-HPV-7 is derived from epithelial cells cultured from normal tissue from the peripheral zone of the prostate. The cells are transformed by transfection with
30 HPV18. Immunocytochemical analysis shows expression of keratins 5 and 8 and also the early region 6 (E6) oncoprotein of HPV. The cells are negative for prostate specific antigen (PSA). CA-HPV-10 is derived from cells from a 63-year-old male with prostatic adenocarcinoma of Gleason Grade 4/4. The cells are transformed by transfection with HPV18 DNA. Immunocytochemical analysis shows expression of keratins 5 and 8 and also the early region 6 (E6) oncoprotein of HPV. The cells are

negative for prostate specific antigen (PSA). DU 145 is a prostate carcinoma cell line isolated from a metastatic site in the brain of 69-year old male with widespread metastatic prostate carcinoma. DU 145 has no detectable sensitivity to hormones; forms colonies in semi-solid medium; is only weakly positive for acid phosphatase; and cells are negative for prostate specific antigen (PSA). LNCaP is a prostate carcinoma cell line isolated from a lymph node biopsy of a 50-year-old male with metastatic prostate carcinoma. LNCaP cells express prostate specific antigens, produce prostatic acid phosphatase, and express androgen receptors. MDAPCa2b is a prostate adenocarcinoma cell line isolated from a metastatic site in the bone of a 63-year-old male. MDAPCa2b expresses prostate specific antigen (PSA) and androgen receptor, grows in vitro and in vivo, and is androgen sensitive. PC-3 is a prostate adenocarcinoma cell line that was isolated from a metastatic site in the bone of a 62-year-old male with grade IV prostate adenocarcinoma.

Starved: cells were grown in basal media in the absence of growth factors and hormones. PrEGM: cells grown under optimal growth conditions, in the presence of growth factors and nutrients. TCH: cells grown in defined serum-free TCH medium. In a further example, expression of SEQ ID NO:54 was decreased 2.2-fold to 10.9-fold in four (DU145, LNCaP, MDAPCa2b, and PC-3) of five cell lines tested (described above) when compared with PrEC or PZ-HPV-7 cells grown in starved, PrECM, or TCH media. Therefore, SEQ ID NO:54 is useful as a diagnostic marker or as a potential therapeutic target for prostate cancer.

SEQ ID NO:56 and SEQ ID NO:64 showed differential expression associated with lung cancer. SEQ ID NO:56 and SEQ ID NO:64 showed at least a two-fold decrease in expression in lung tissue from three patients with lung adenocarcinoma and four patients with squamous cell carcinoma compared to matched microscopically normal tissue from the same donors as determined by microarray analysis. Moderately differentiated adenocarcinoma of the right lung was compared to grossly uninvolved lung tissue from a 60 year-old donor (Huntsman Cancer Institute, Salt Lake City, UT). Grossly uninvolved lung tissue from a 66 year-old female was compared to lung adenocarcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). Grossly uninvolved lung tissue from a 71 year-old female was compared to lung adenocarcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Grossly uninvolved lung tissue from a 73 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Grossly uninvolved lung tissue from a 68 year-old female was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Grossly uninvolved lung tissue from a 66 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung

Cancer Research). Grossly uninvolved lung tissue from a 73 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Therefore, SEQ ID NO:56 and SEQ ID NO:64 are useful in disease staging and diagnostic assays for cell proliferative disorders, including lung cancer.

5 SEQ ID NO:58 and SEQ ID NO:59 showed differential expression associated with breast cancer as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma cell lines representing different stages of tumor progression. The cell lines compared included: a) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass
10 isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e)
15 MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MDA-MB-468, a breast adeoncarcinoma cell line isolated from the pleural effusion of a 51-year old female with metastatic adenocarcinoma of the breast, g) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, and h) HMEC, a primary breast epithelial cell line isolated from a normal
20 donor. The expression levels of SEQ ID NO:58 and SEQ ID NO:59 were at least two-fold lower in the BT-474, BT-483, Hs578T, and MCF7 breast carcinoma cell lines compared to the HMEC cell line. In addition, SEQ ID NO:58 and SEQ ID NO:59 showed at least a two-fold decrease in expression in MCF7 and BT-20 cells treated with 50 ng/ml epidermal growth factor compared to untreated cells. Epidermal growth factor is highly expressed in breast carcinoma cells. Epidermal
25 growth factor promotes proliferation and differentiation of mesenchymal and epithelial cells, angiogenesis, and tumor progression. It is a mitogen for fibroblasts, epithelial and endothelial cells. In an alternative example, SEQ ID NO:58 and SEQ ID NO:59 showed differential expression associated with prostate cancer as determined by microarray analysis. The expression of SEQ ID NO:58 and SEQ ID NO:59 showed at least a two-fold decrease in prostate PC-3, LNCaP and DU-145
30 carcinoma cells compared to prostate PrEC epithelial cells. The PC-3 cell line was isolated from a 62-year old male with grade IV prostate adenocarcinoma. The LNCaP cell line was isolated from a lymph node biopsy of a 50-year old male with metastatic prostate carcinoma. The DU-145 cell line was isolated from a 69-year old male with widespread metastatic prostate carcionoma. The PrEC cell line is a prostate epithelial cell line isolated from a normal donor. Therefore SEQ ID NO:58 and SEQ

ID NO:59 are useful in diagnostic assays and disease staging assays for cell proliferative disorders, including breast cancer and prostate cancer.

In another example, the expression of SEQ ID NO:61 was decreased by at least two-fold in human preadipocytes from obese and normal donors treated with a differentiation inducing medium when compared to non-treated preadipocytes from the same donors. The normal human primary subcutaneous preadipocytes were isolated from adipose tissue of a 28-year-old healthy female with a body mass index (BMI) of 23.59. The obese human primary subcutaneous preadipocytes were isolated from adipose tissue of a 40-year-old healthy female with a body mass index (BMI) of 32.47. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in the differentiation medium containing the active components, PPAR- γ agonist and human insulin. Human preadipocytes were treated with human insulin and PPAR- γ agonist for three days and subsequently were switched to medium containing insulin for 24 hours, 48 hours, 4 days, 8 days or 15 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. Between 80% and 90% of the preadipocytes finally differentiated to adipocytes as observed under a phase contrast microscope. The experiments showed that at three out of five time points (4, 8 and 15 days), the expression of SEQ ID NO:61 was decreased by at least 2 fold in adipocytes from an obese donor, and at 15 days, the expression of SEQ ID NO:61 was decreased by at least 2 fold in human adipocytes from a normal donor. Therefore, SEQ ID NO:61 is useful for the diagnosis, prognosis, or treatment of disorders, including diabetes, obesity, hypertension, and atherosclerosis.

In an alternative example, the expression of SEQ ID NO:65 showed at least a two-fold decrease in proliferating human artery cells (ASMCs) or fibroblasts compared to growth inhibited ASMC or fibroblasts. Human coronary ASMCs and lung fibroblasts were grown and harvested in both a proliferative and growth inhibited state. The process of formation of atherosclerotic lesions begins with a protective response to insults to the endothelium and smooth muscle cells of the wall of the artery. This eventually becomes an excessive inflammatory-fibroproliferative response. The resulting lesions are composed of layers of macrophages and smooth muscle cells over a lipid core and covered with a cap of connective tissue. Smooth muscle cells are believed to switch from a contractile to a synthetic phenotype in which quiescent cells become migratory and proliferative and capable of synthesizing different extracellular matrix components. Therefore, SEQ ID NO:65 is useful for the diagnosis, prognosis, or treatment of cardiovascular disorders, including atherosclerosis.

In an alternative example, SEQ ID NO:65 showed differential expression associated with breast cancer as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of breast carcinoma cell

lines representing different stages of tumor progression. The cell lines compared included: a) BT-20, a breast carcinoma cell line derived *in vitro* from the cells emigrating out of thin slices of tumor mass isolated from a 74-year-old female, b) BT-474, a breast ductal carcinoma cell line that was isolated from a solid, invasive ductal carcinoma of the breast obtained from a 60-year-old woman, c) BT-483, a breast ductal carcinoma cell line that was isolated from a papillary invasive ductal tumor obtained from a 23-year-old normal, menstruating, parous female with a family history of breast cancer, d) Hs 578T, a breast ductal carcinoma cell line isolated from a 74-year-old female with breast carcinoma, e) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, f) MDA-MB-468, a breast adeoncarcinoma cell line isolated from the pleural effusion of a 51-year old female with metastatic adenocarcinoma of the breast, g) MCF-10A, a breast mammary gland (luminal ductal characteristics) cell line isolated from a 36-year-old woman with fibrocystic breast disease, h) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year old female with an infiltrating ductal carcinoma of the breast, i) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year old female, j) HMEC, a primary breast epithelial cell line isolated from a normal donor. The expression levels of SEQ ID NO:65 were at least two-fold lower in the BT-20, BT-474, BT-483, MCF-10A, MCF7, T-47D, and Sk-BR-3 breast carcinoma cell lines compared to the HMEC cell line.

In addition, SEQ ID NO:65 showed at least a two-fold increase in expression in MCF-10A cells treated with 1 ng/ml transforming growth factor- β (TGF- β) for 24 or 36 hours compared to untreated cells. In an alternative example, SEQ ID NO:65 showed differential expression in colon tissue from patients with colon polyps or colon cancer compared to matched microscopically normal tissue from the same donors as determined by microarray analysis. The expression of SEQ ID NO:65 was decreased at least two-fold in patients with colon polyps and adenocarcinoma and increased at least two-fold in a patient with colon sarcoma. Colon tissue from two colon polyps was compared to grossly uninvolved colon tissue from a 23 year-old donor diagnosed with familial polyposis syndrome and colon adenocarcinoma (Huntsman Cancer Institute). Colon adenocarcinoma tissue from a 38 year-old male donor was compared to normal colon tissue from the same donor (Huntsman Cancer Institute). Colon tissue from a sigmoid colon tumor originating from a metastatic gastric sarcoma from a 48 year-old female donor was compared to grossly uninvolved sigmoid colon tissue from the same donor (Huntsman Cancer Institute).

In an alternative example, SEQ ID NO:65 showed differential expression associated with prostate cancer as determined by microarray analysis. The expression of SEQ ID NO:20 showed at least a two-fold decrease in prostate LNCaP and DU-145 carcinoma cells compared to prostate nontumorigenic PZ-HPV-7 epithelial cells. The expression of SEQ ID NO:65 showed at least a two-

fold decrease in prostate MDAPCa2b carcinoma cells compared to prostate PrEC epithelial cells. The PC-3 cell line was isolated from a 62-year old male with grade IV prostate adenocarcinoma. The LNCaP cell line was isolated from a lymph node biopsy of a 50-year old male with metastatic prostate carcinoma. The DU-145 cell line was isolated from a 69-year old male with widespread metastatic prostate carcinoma. The PZ-HPV-7 cell line was derived from epithelial cells cultured from normal tissue from the peripheral zone of the prostate. MDAPCa2b is a prostate adenocarcinoma cell line that was isolated from the bone metastatic site of a 63-year-old male. The PrEC cell line is a prostate epithelial cell line isolated from a normal donor. In addition, SEQ ID NO:65 showed at least a two-fold increase in expression in PrEC cells treated with 5 ng/ml TGF- β for 8, 14, 24, or 38 hours.

In an alternative example, SEQ ID NO:65 showed at least a two-fold decrease in expression associated with endometrial cancer as determined by microarray analysis. Endometrial adenocarcinoma tissue was compared to grossly uninvolved endometrial tissue from a 72 year-old female donor (Huntsman Cancer Institute). In an alternative example, SEQ ID NO:65 showed differential expression associated with lung cancer. SEQ ID NO:65 showed at least a two-fold decrease in expression in lung tissue from a patient with lung adenocarcinoma and at least a two-fold increase in expression in four patients with squamous cell carcinoma compared to matched microscopically normal tissue from the same donors as determined by microarray analysis. Moderately differentiated adenocarcinoma of the right lung was compared to grossly uninvolved lung tissue from a 60 year-old donor (Huntsman Cancer Institute). Grossly uninvolved lung tissue from a 75 year-old female was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Grossly uninvolved lung tissue from a 73 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Grossly uninvolved lung tissue from a 66 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Grossly uninvolved lung tissue from a 73 year-old male was compared to lung squamous cell carcinoma tissue from the same donor (Roy Castle International Centre for Lung Cancer Research). Therefore SEQ ID NO:65 is useful in diagnostic assays and disease staging assays for cell proliferative disorders, including breast cancer, colon cancer, prostate cancer, endometrial cancer, and lung cancer.

In another example, the expression of SEQ ID NO:65 was decreased by at least two-fold in ECV304 immortalized endothelial cells treated with cytokines, including TNF- α , IFN- γ , or IL-1 β . ECV304 cells were grown to 85% confluency and then treated with 10 ng/ml IFN- γ for 24 hours. The ECV304 cells were next treated with 10 ng/ml TNF- α for 0.67, 1, 4, 8, 24, 48, and 72 hours. The cells treated with cytokines were compared to untreated ECV304 cells collected at 85% confluency.

The expression of SEQ ID NO:65 was decreased by at least two-fold in ECV304 cells that were stimulated with 10 ng/ml TNF- α for 4, 8 or 24 hours when compared to untreated ECV304 cells. In another experiment, ECV304 cells were grown to 85% confluency and then treated with 10 ng/ml IL-1 β and 10 ng/ml TNF- α for 0.67, 1, 4, 8, 24, 48, and 72 hours. SEQ ID NO:65 showed at least a two-fold decrease in expression in ECV304 cells that were stimulated with IL-1 β and TNF- α compared to untreated ECV304 cells at all of the time points. In another experiment, ECV304 cells were grown to 85% confluency and then treated with 10 ng/ml TNF- α for 0.67, 1, 4, 8, 24, 48, and 72 hours. SEQ ID NO:65 showed at least a two-fold decrease in expression in ECV304 cells that were stimulated TNF- α for 2, 4, 8 or 24 hours compared to untreated ECV304 cells. In another experiment, ECV304 cells were grown to 85% confluency and then treated with 0.01, 0.03, 0.1, 0.3, 1, or 3 ng/ml TNF- α for 0, 1, 2, 8, or 24 hours. SEQ ID NO:65 showed at least a two-fold decrease in expression in ECV304 cells that were stimulated with 1, 3, or 10 ng/ml TNF- α for 2, 8 or 24 hours compared to untreated ECV304 cells. The ECV304 cell line is derived from the endothelium of the human umbilical vein. This cell model has been extensively used as an experimental model for investigating *in vitro* the role of endothelium in human vascular biology. IFN- γ is a cytokine, produced primarily by T-lymphocytes and natural killer cells, that exerts antiproliferative, immunoregulatory, and proinflammatory activities, and plays a role in host defense mechanisms. IFN- γ induces the production of cytokines, upregulates the expression of class I and II MHC antigens, and Fc receptor and leukocyte adhesion molecules. It modulates macrophage effector functions, influences isotype switching and potentiates the secretion of immunoglobulins by B cells. IFN- γ also augments TH1 cell expansion and may be required for TH1 cell differentiation. IL-1 β is a cytokine associated with acute inflammatory responses and is generally considered the prototypical pro-inflammatory cytokine. TNF- α is a pleiotropic cytokine that plays a central role in mediating the inflammatory response through activation of multiple signal transduction pathways. TNF- α is produced by activated lymphocytes, macrophages, and other white blood cells and can activate endothelial cells. Monitoring the endothelial cells' response to TNF- α at the level of mRNA expression can provide information necessary for better understanding of both TNF- α signaling pathways and endothelial cell biology. Therefore, SEQ ID NO:65 is useful in diagnosis, prognosis, or treatment of inflammatory disorders and endothelial disorders, including disorders of vascular tone regulation, coagulation, thrombosis, and atherosclerosis.

In another example, the expression of SEQ ID NO:65 was increased by at least four-fold in human preadipocytes from an overweight donor treated with a differentiation inducing medium when compared to non-treated preadipocytes from the same donor. The human primary subcutaneous preadipocytes were isolated from adipose tissue of an overweight 36-year-old healthy female with a

body mass index (BMI) of 27.7. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in the differentiation medium containing the active components, PPAR- γ agonist and human insulin. Human preadipocytes were treated with human insulin and PPAR- γ agonist for three days and subsequently were switched to medium containing insulin for 5, 9 and 12 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. More than 60% of the preadipocytes differentiated to adipocytes after 15 days in culture as observed under a phase contrast microscope. The expression of SEQ ID NO:65 increased by at least 4-fold in adipocytes from an overweight donor at 1.1, 1.7, and 2.1 weeks. In another experiment, the expression of SEQ ID NO:65 was increased by at least two-fold in human preadipocytes from normal donor treated with a differentiation inducing medium when compared to non-treated preadipocytes from the same donor. The human primary subcutaneous preadipocytes were isolated from adipose tissue of a normal 28-year-old healthy female with a body mass index (BMI) of 23.59. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in the differentiation medium containing the active components, PPAR- γ agonist and human insulin. Human preadipocytes were treated with human insulin and PPAR- γ agonist for three days and subsequently were switched to medium containing insulin for 1, 2, 3, 8, 15, and 20 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. Between 80% and 90% of the preadipocytes differentiated to adipocytes in culture as observed under a phase contrast microscope. The expression of SEQ ID NO:65 increased by at least two-fold in adipocytes from a normal donor at 1, 2, 8, and 15 days. In another experiment, the expression of SEQ ID NO:65 was decreased by at least two-fold in human preadipocytes from an obese donor treated with a differentiation inducing medium when compared to non-treated preadipocytes from the same donor. The human primary subcutaneous preadipocytes were isolated from adipose tissue of a n obese 40-year-old healthy female with a body mass index (BMI) of 32.47. The preadipocytes were cultured and induced to differentiate into adipocytes by culturing them in the differentiation medium containing the active components, PPAR- γ agonist and human insulin. Human preadipocytes were treated with human insulin and PPAR- γ agonist for three days and subsequently were switched to medium containing insulin for 1, 2, 3, 4, 8, 15, and 20 days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. Between 80% and 90% of the preadipocytes differentiated to adipocytes in culture as observed under a phase contrast microscope. The expression of SEQ ID NO:65 decreased by at least two-fold in adipocytes from an obese donor at 1, 2, and 4 days. Therefore, SEQ ID NO:65 is useful for the diagnosis, prognosis, or treatment of

disorders, including diabetes, obesity, hypertension, and atherosclerosis.

For example, SEQ ID NO:74 showed differential expression associated with breast cancer, as determined by microarray analysis. The gene expression profile of a nonmalignant mammary epithelial cell line was compared to the gene expression profiles of various breast carcinoma lines at different stages of tumor progression. Cell lines compared included: a) HMEC, a primary breast epithelial cell line isolated from a normal donor, b) MCF-10A, a breast mammary gland cell line isolated from a 36-year-old woman with fibrocystic breast disease, c) MCF7, a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female, d) T-47D, a breast carcinoma cell line isolated from a pleural effusion obtained from a 54-year-old female with an infiltrating ductal carcinoma of the breast, e) Sk-BR-3, a breast adenocarcinoma cell line isolated from a malignant pleural effusion of a 43-year-old female, f) BT-20, a breast carcinoma cell line derived *in vitro* from cells emigrating out of thin slices of the tumor mass isolated from a 74-year-old female, g) MDA-mb-231, a breast tumor cell line isolated from the pleural effusion of a 51-year-old female, and h) MDA-mb-435S, a spindle-shaped strain that evolved from the parent line (435) isolated by R. Cailleau from a pleural effusion of a 31-year-old female with metastatic, ductal adenocarcinoma of the breast. The expression of SEQ ID NO:74 was reduced by more than four-fold in BT20 and T-47D cells, by nearly 8-fold in MCF7 cells, and by over 10-fold in Sk-BR-3 cells, as compared to HMEC cells. Therefore, SEQ ID NO:74 is useful for monitoring treatment of, and diagnostic assays for breast cancer.

In another example, SEQ ID NO:73 and SEQ ID NO:74 showed differential expression associated with colon cancer and polyp development in the colon, respectively, as determined by microarray analysis. For SEQ ID NO:73, normal colon tissue from a 56-year-old female diagnosed with poorly differentiated metastatic adenocarcinoma of possible ovarian origin and a clinical history of recurrent cecal mass was compared to colon tumor tissue from the same donor (Huntsman Cancer Institute, Salt Lake City, UT) by competitive hybridization. SEQ ID NO:73 showed at least a two-fold increase in expression in tumor samples when compared to normal tissue samples. When gene expression profiles were compared between normal colon tissue from a 23-year-old donor diagnosed with FAP (familial polyposis syndrome) and tissue from donor polyp, SEQ ID NO:74 showed at least a 3-fold decrease in expression in the polyp-derived tissue versus normal colon tissue from the same donor. Thus, SEQ ID NO:73 and SEQ ID NO:74 are useful in monitoring treatment of, and diagnostic assays for colon polyps and colon cancer.

In another example, SEQ ID NO:74 displays differential expression in a number of lung carcinoma samples, as determined by microarray analysis. Moderately differentiated adenocarcinoma tissue of the right lung was compared to grossly uninvolved lung tissue from a 60-year-old donor

(Huntsman Cancer Institute). SEQ ID NO:74 showed a decrease in expression of at least two-fold in the tumor tissue, when compared with the normal lung tissue. Also, the expression profile of normal lung tissue was compared to the expression profile of squamous cell carcinoma tissue from the lung of several matched donor pairs, including a 75-year-old female donor, a 73-year-old female donor, a 66-year-old male donor, and a 73-year-old male donor (Roy Castle International Centre for Lung Cancer Research, Liverpool, UK). SEQ ID NO:74 displayed an increase in expression in the tumor samples versus the normal tissue samples of between two- to 4.5-fold when examined by microarray analysis. Therefore, SEQ ID NO:74 will be useful in monitoring treatment of, and diagnostic assays for lung cancer.

In yet another example, SEQ ID NO:73 and SEQ ID NO:74 were shown to have differential expression associated with osteosarcoma tumors and metastases to the lung, as determined by microarray analysis. Messenger RNA from normal human osteoblasts was compared with mRNA from biopsy specimens, osteosarcoma tissues, primary cultures, or metastasized tissues. A normal osteoblast primary culture, NHOst 5488, was chosen as the reference in the initial experiments. One basic set of experiments is defined as the comparison of mRNA from biopsy specimen with that of definitive surgical specimen from the same patient. Extended study of this basic set includes mRNA from primary cell cultures of the definitive surgical specimen, muscle, or cartilage tissue from the same patient. Biopsy specimens, definitive surgical specimens, or lung metastatic tissues from different individuals were also included to reveal individual variability. SEQ ID NO:73 expression in osteosarcoma samples from several donors was decreased when compared to the gene expression profile of the normal osteoblast reference cell line. On average, SEQ ID NO:73 gene expression decreased by at least two-fold in each set of tissue samples or cell lines derived from either a spindle cell or a chondroblastic osteosarcoma patient, when compared to normal osteoblasts. In another chondroblastic osteosarcoma tumor sample, SEQ ID NO:73 expression was over 9-fold reduced when compared to the gene expression profile of normal osteoblasts. In a lung tumor sample, representing a metastatic lesion from an osteosarcoma, SEQ ID NO:73 gene expression was reduced about three-fold in comparison to its expression in normal osteoblasts, while SEQ ID NO:74 gene expression was actually increased over 4-fold in the metastatic tissue sample, relative to normal osteoblast expression levels. Therefore, SEQ ID NO:73 and SEQ ID NO:74 will be useful for monitoring of, and diagnostic assays for osteosarcomas.

In another example, SEQ ID NO:74 gene expression was monitored during the differentiation of human mesenchymal stem cells (HMSCs), which are found in the bone marrow, into osteoblasts over a time course. After 3.1 weeks of culture in differentiation-inducing media, expression of SEQ ID NO:74 in the HMSCs was increased at least two-fold over untreated HMSCs. SEQ ID NO:74 is

thus also useful for monitoring differentiation of stem cells into osteoblasts.

In a further example, SEQ ID NO:74 showed differential expression associated with obesity. In this series of experiments, human primary subcutaneous preadipocytes were isolated from adipose tissue of donors with a range of body mass index (BMI) measurements, which indicate whether a person is considered of healthy weight ($BMI < 25$), overweight ($25 < BMI < 30$), or obese ($BMI > 30$). Peroxisome-proliferator-activated receptor gamma ($PPAR\gamma$), a member of the nuclear hormone receptor superfamily, is a potent inducer of adipocyte differentiation when ectopically expressed in cultured preadipocytes. Human primary subcutaneous preadipocytes isolated from the adipose tissue of an overweight donor ($BMI = 27.7$) were induced to differentiate into adipocytes by culturing them in differentiation medium containing active components such as $PPAR\gamma$ agonist and human insulin for three days. The cells were subsequently switched to medium containing insulin only for 5, 9, and 12 more days before the cells were collected for analysis. Differentiated adipocytes were compared to untreated preadipocytes maintained in culture in the absence of inducing agents. At least 60% of the preadipocytes finally differentiated to adipocytes as observed under phase contrast microscope. In these cells, SEQ ID NO:74 gene expression increased at least two-fold after 8 days, and then further increased over 4-fold after 12 and 15 days, when compared to untreated preadipocytes. In a similar experiment, preadipocytes isolated in the same manner from a healthy donor ($BMI = 23.59$) showed an increase of at least two-fold in SEQ ID NO:74 expression upon treatment with differentiation media, when compared to the gene expression profile from untreated cells from the same donor. However, when preadipocytes from an obese donor ($BMI = 32.47$) were examined after treatment with differentiation media, expression of SEQ ID NO:74 actually decreased at least two-fold, when compared to untreated preadipocytes from this donor. Thus, SEQ ID NO:74 is useful for the diagnosis, prognosis, or treatment of disorders such as obesity.

In a final example, SEQ ID NO:73 showed differential expression in human umbilical vein endothelial cells (HUVEC), as measured by microarray analysis. HUVEC cells have been extensively used to study the functional biology of human endothelial cells *in vitro*. Activation of the vascular endothelium is a central event in normal and pathological inflammatory responses. $TNF\alpha$ and $IL-1\beta$ are two important pro-inflammatory cytokines that initiate activation of vascular endothelium. Cultures were transfected with lipofectin with or without an antisense oligonucleotide that targets the degradation of WSB-1 mRNA (which did not affect SEQ ID NO:73 expression levels), then treated with recombinant human $TNF\alpha$ and $IL-1\beta$ at 10 ng/ml each for 24 hours or were left untreated. Compared to the gene expression profile of untreated HUVECs, SEQ ID NO:73 levels increased at least three-fold upon $TNF\alpha$ and $IL-1\beta$ treatment. Therefore, SEQ ID NO:73 is useful in diagnosis of, and monitoring treatment for, autoimmune/inflammatory disorders.

XII. Complementary Polynucleotides

Sequences complementary to the CADECM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CADECM. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same
5 procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CADECM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CADECM-encoding
10 transcript.

XIII. Expression of CADECM

Expression and purification of CADECM is achieved using bacterial or virus-based expression systems. For expression of CADECM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of
15 cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express CADECM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CADECM in eukaryotic cells is achieved by infecting
20 insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CADECM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to
25 infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, CADECM is synthesized as a fusion protein with, e.g.,
30 glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from CADECM

at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified CADECM obtained by these methods can be used directly in the assays shown in Examples XVII and XVIII, where applicable.

XIV. Functional Assays

CADECM function is assessed by expressing the sequences encoding CADECM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; *Flow Cytometry*, Oxford, New York NY).

The influence of CADECM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CADECM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake

Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding CADECM and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of CADECM Specific Antibodies

CADECM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the CADECM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CADECM activity by, for example, binding the peptide or CADECM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring CADECM Using Specific Antibodies

Naturally occurring or recombinant CADECM is substantially purified by immunoaffinity chromatography using antibodies specific for CADECM. An immunoaffinity column is constructed by covalently coupling anti-CADECM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CADECM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CADECM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CADECM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CADECM is collected.

XVII. Identification of Molecules Which Interact with CADECM

CADECM, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539). Candidate molecules

previously arrayed in the wells of a multi-well plate are incubated with the labeled CADECM, washed, and any wells with labeled CADECM complex are assayed. Data obtained using different concentrations of CADECM are used to calculate values for the number, affinity, and association of CADECM with the candidate molecules.

5 Alternatively, molecules interacting with CADECM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

 CADECM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all
10 interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of CADECM Activity

 An assay for CADECM activity measures the expression of CADECM on the cell surface. cDNA encoding CADECM is transfected into a non-leukocytic cell line. Cell surface proteins are
15 labeled with biotin (de la Fuente, M.A. et al. (1997) Blood 90:2398-2405). Immunoprecipitations are performed using CADECM-specific antibodies, and immunoprecipitated samples are analyzed using SDS-PAGE and immunoblotting techniques. The ratio of labeled immunoprecipitant to unlabeled immunoprecipitant is proportional to the amount of CADECM expressed on the cell surface.

 Alternatively, an assay for CADECM activity measures the amount of cell aggregation
20 induced by overexpression of CADECM. In this assay, cultured cells such as NIH3T3 are transfected with cDNA encoding CADECM contained within a suitable mammalian expression vector under control of a strong promoter. Cotransfection with cDNA encoding a fluorescent marker protein, such as Green Fluorescent Protein (CLONTECH), is useful for identifying stable transfectants. The amount of cell agglutination, or clumping, associated with transfected cells is compared with that
25 associated with untransfected cells. The amount of cell agglutination is a direct measure of CADECM activity.

 Alternatively, an assay for CADECM activity measures the disruption of cytoskeletal filament networks upon overexpression of CADECM in cultured cell lines (Rezniczek, G. A. et al. (1998) J. Cell Biol. 141:209-225). cDNA encoding CADECM is subcloned into a mammalian
30 expression vector that drives high levels of cDNA expression. This construct is transfected into cultured cells, such as rat kangaroo PtK2 or rat bladder carcinoma 804G cells. Actin filaments and intermediate filaments such as keratin and vimentin are visualized by immunofluorescence microscopy using antibodies and techniques well known in the art. The configuration and abundance of cytoskeletal filaments can be assessed and quantified using confocal imaging techniques. In

particular, the bundling and collapse of cytoskeletal filament networks is indicative of CADECM activity.

Alternatively, cell adhesion activity in CADECM is measured in a 96-well plate in which wells are first coated with CADECM by adding solutions of CADECM of varying concentrations to the wells. Excess CADECM is washed off with saline, and the wells incubated with a solution of 1% bovine serum albumin to block non-specific cell binding. Aliquots of a cell suspension of a suitable cell type are then added to the wells and incubated for a period of time at 37 °C. Non-adherent cells are washed off with saline and the cells stained with a suitable cell stain such as Coomassie blue. The intensity of staining is measured using a variable wavelength multi-well plate reader and compared to a standard curve to determine the number of cells adhering to the CADECM coated plates. The degree of cell staining is proportional to the cell adhesion activity of CADECM in the sample.

Alternatively, CADECM activity may be demonstrated as the ability to interact with its associated LMW GTPase in an in vitro binding assay. The candidate LMW GTPases are expressed as fusion proteins with glutathione S-transferase (GST), and purified by affinity chromatography on glutathione-Sepharose. The LMW GTPases are loaded with GDP by incubating 20 mM Tris buffer, pH 8.0, containing 100 mM NaCl, 2 mM EDTA, 5 mM MgCl₂, 0.2 mM DTT, 100 μM AMP-PNP and 10 μM GDP at 30°C for 20 minutes. CADECM is expressed as a FLAG fusion protein in a baculovirus system. Extracts of these baculovirus cells containing CADECM-FLAG fusion proteins are precleared with GST beads, then incubated with GST-GTPase fusion proteins. The complexes formed are precipitated by glutathione-Sepharose and separated by SDS-polyacrylamide gel electrophoresis. The separated proteins are blotted onto nitrocellulose membranes and probed with commercially available anti-FLAG antibodies. CADECM activity is proportional to the amount of CADECM-FLAG fusion protein detected in the complex.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be

defined by the following claims and their equivalents.

Table 1

| Incyte Project ID | Polypeptide SEQ ID NO: | Incyte Polypeptide ID | Polynucleotide SEQ ID NO: | Incyte Polynucleotide ID |
|-------------------|---------------------------|--------------------------|------------------------------|--------------------------------|
| 7504868 | 1 | 7504868CD1 | 38 | 7504868CB1 |
| 7504930 | 2 | 7504930CD1 | 39 | 7504930CB1 |
| 6610456 | 3 | 6610456CD1 | 40 | 6610456CB1 |
| 7503573 | 4 | 7503573CD1 | 41 | 7503573CB1 |
| 7505057 | 5 | 7505057CD1 | 42 | 7505057CB1 |
| 90116002 | 6 | 90116002CD1 | 43 | 90116002CB1 |
| 39283 | 7 | 039283CD1 | 44 | 039283CB1 |
| 7505082 | 8 | 7505082CD1 | 45 | 7505082CB1 |
| 7505139 | 9 | 7505139CD1 | 46 | 7505139CB1 |
| 7505234 | 10 | 7505234CD1 | 47 | 7505234CB1 |
| 7500227 | 11 | 7500227CD1 | 48 | 7500227CB1 |
| 7503676 | 12 | 7503676CD1 | 49 | 7503676CB1 |
| 7503606 | 13 | 7503606CD1 | 50 | 7503606CB1 |
| 7500216 | 14 | 7500216CD1 | 51 | 7500216CB1 |
| 7099880 | 15 | 7099880CD1 | 52 | 7099880CB1 |
| 871513 | 16 | 871513CD1 | 53 | 871513CB1 |
| 8057640 | 17 | 8057640CD1 | 54 | 8057640CB1 |
| 7505913 | 18 | 7505913CD1 | 55 | 7505913CB1 |
| 7510292 | 19 | 7510292CD1 | 56 | 7510292CB1 |
| 7504669 | 20 | 7504669CD1 | 57 | 7504669CB1 |
| 7509266 | 21 | 7509266CD1 | 58 | 7509266CB1 |
| 7509288 | 22 | 7509288CD1 | 59 | 7509288CB1 |
| 7510212 | 23 | 7510212CD1 | 60 | 7510212CB1 |
| 7510504 | 24 | 7510504CD1 | 61 | 7510504CB1 |
| 7510587 | 25 | 7510587CD1 | 62 | 7510587CB1 |
| 7510684 | 26 | 7510684CD1 | 63 | 7510684CB1 |
| 7510697 | 27 | 7510697CD1 | 64 | 7510697CB1 |
| 7761337 | 28 | 7761337CD1 | 65 | 7761337CB1 |
| 7503666 | 29 | 7503666CD1 | 66 | 7503666CB1 |
| 7503668 | 30 | 7503668CD1 | 67 | 7503668CB1 |
| 7503672 | 31 | 7503672CD1 | 68 | 7503672CB1 |
| 6039650 | 32 | 6039650CD1 | 69 | 6039650CB1 |
| 7509919 | 33 | 7509919CD1 | 70 | 7509919CB1 |
| 7510758 | 34 | 7510758CD1 | 71 | 7510758CB1 |
| 7510063 | 35 | 7510063CD1 | 72 | 7510063CB1 |
| 7510135 | 36 | 7510135CD1 | 73 | 7510135CB1 |
| 7505011 | 37 | 7505011CD1 | 74 | 7505011CB1 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 1 | 7504868CD1 | g2177172 | 1.1E-169 | [Homo sapiens] mucosal addressin cell adhesion molecule-1 Leung, E. et al. (1997) Genomic organization, chromosomal mapping, and analysis of the 5' promoter region of the human MADCAM-1 gene. Immunogenetics 46:111-119 |
| | | 305407 MADCA M1 | 8.0E-92 | [Homo sapiens] [Adhesin/agglutinin; Receptor signalling] Mucosal addressin cell adhesion molecule 1, an integrin ligand of the immunoglobulin family, involved in lymphocyte homing to Peyer's patch high endothelial venules, upregulated in ulcerative colitis and Crohn's disease |
| | | | | Souza, H. S. et al. (1999) Expression of lymphocyte-endothelial receptor-ligand pairs, alpha4beta7/MADCAM-1 and OX40/OX40 ligand in the colon and jejunum of patients with inflammatory bowel disease. Gut 45: 856-863 |
| | | 717204 1bqs_A | 7.1E-62 | [Protein Data Bank] Mucosal Addressin Cell Adhesion Molecule-1 |
| | | 587143 Madcam1 | 6.8E-50 | [Mus musculus] [Adhesin/agglutinin] [Plasma membrane] Mucosal vascular addressin, a member of the immunoglobulin superfamily and a cell adhesion molecule for lymphocytes that has a role in homing lymphocytes to high endothelial venules in lymph nodes |
| | | 609753 Madcam1 | 4.2E-46 | [Rattus norvegicus] [Plasma membrane] Member of the immunoglobulin superfamily, has strong similarity to murine Madcam1, which is a cell adhesion molecule for lymphocytes that has a role in homing lymphocytes to high endothelial venules in lymph nodes |
| | | 336506 MUC2 | 1.4E-15 | [Homo sapiens] [Secretory vesicles; Cytoplasmic; E14Extracellular (excluding cell wall)] Mucin 2, intestinal glycoprotein secreted by goblet cells; expression is altered in diseased gastric, colonic and intestinal neoplastic cells; induced by Pseudomonas aeruginosa leading to airway mucus obstruction in patients with cysticfibrosis |
| 2 | 7504930CD1 | g2078518 | 0.0 | [Homo sapiens] neogenin Vielmetter, J. et al. (1997) Molecular characterization of human neogenin, a DCC-related protein, and the mapping of its gene (NEO1) to chromosomal position 15q22.3-q23. Genomics 41:414-421 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 2 | cont | 336588 NEO1 | 0.0 | [Homo sapiens] [Adhesin/agglutinin] [Plasma membrane] Neogenin, a member of the N-CAM family of cell adhesion molecules, may regulate differentiation and/or cell migration and may play a role in neurogenesis Meyerhardt, J. A. et al. (1997) Identification and characterization of neogenin, a DCC-related gene. <i>Oncogene</i> 14:1129-1136 |
| | | 330916 Neo1 | 0.0 | [Rattus norvegicus] [Adhesin/agglutinin; Receptor (signalling)] [Plasma membrane] Neogenin, may bind netrin-1, may regulate spinal cord differentiation and cell migration |
| | | 368654 Neo1 | 0.0 | [Mus musculus] [Adhesin/agglutinin] [Plasma membrane] Neogenin, a member of the N-CAM family of cell adhesion molecules, may regulate differentiation and/or cell migration and may play a role in neurogenesis |
| | | 330914 Dcc | 0.0 | [Rattus norvegicus] [Receptor (signalling)] [Plasma membrane] Deleted in colorectal cancer, netrin-1 receptor, member of the immunoglobulin-CAM family which mediates axon guidance and regulates apoptosis; human DCC is deleted in colorectal cancers |
| | | 429560 Dcc | 0.0 | [Mus musculus] [Receptor (signalling)] [Plasma membrane] Deleted in colorectal cancer, netrin-1 receptor, member of the immunoglobulin-CAM family which is involved in axon guidance and has a role in regulating apoptosis; human DCC is deleted in colorectal cancers |
| 3 | 6610456CD1 | g4322670 | 9.6E-11 | [Homo sapiens] dentin phosphorin Gu, K. et al. (1998) Human dentin phosphorin nucleotide and amino acid sequence. <i>Eur. J. Oral Sci.</i> 106:1043-1047 |
| | | 435702 DSPP | 8.4E-12 | [Homo sapiens] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall) Dentin sialophosphoprotein (phosphoryn), an extracellular matrix protein of dentin in teeth, contains RGD sequence motifs, may function in dentin mineralization; gene mutation is associated with dentinogenesis imperfecta Shields type II Rowe, P. S. et al. (2000) MEPE, a new gene expressed in bone marrow and tumors causing osteomalacia. <i>Genomics</i> 67:54-68 |
| 4 | 7503573CD1 | g3327808 | 0.0 | [Homo sapiens] latent transforming growth factor-beta binding protein 4S Saharinen, J. et al. (1998) Identification and characterization of a new latent transforming growth factor-binding protein, LTBP-4. <i>J. Biol. Chem.</i> 273:18459-18469 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|--|
| 4 | cont | 336298 LTBP4 | 0.0 | [Homo sapiens] [Small molecule-binding protein] [Extracellular matrix (cuticle and basement membrane)] Member of a family of extracellular microfibrillar proteins which also bind transforming growth factor-beta Giltay, R. et al. (1997) Sequence and expression of a novel member (LTBP-4) of the family of latent transforming growth factor-beta binding proteins. FEBS Lett 411:164-168 |
| | | 339488 LTBP2 | 2.4E-252 | [Homo sapiens] [Regulatory subunit; Anchor Protein; Inhibitor or repressor; Small molecule-binding protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Latent transforming growth factor (TGF)-beta binding protein-2, a putative structural component of microfibrils, contains EGF-like repeats and a cysteine-rich region, may be involved in TGF-beta signalling |
| | | 339486 LTBP1 | 2.2E-249 | [Homo sapiens] [Small molecule-binding protein] Latent transforming growth factor-beta binding protein 1, plays a role in latent TGF-beta 1 (TGFB1) assembly, secretion, targeting to the extracellular matrix, and availability; variants may be associated with coronary heart disease |
| | | 619058 Ltbp1 | 2.1E-238 | [Rattus norvegicus] [Inhibitor or repressor] Latent transforming growth factor-beta binding protein 1, may play a role in latent TGF-beta 1 (Tgfb1) assembly, targeting to the extracellular matrix, and availability; variants of human LTBP1 may be associated with coronary heart disease |
| | | 418532 Ltbp2 | 1.7E-236 | [Mus musculus] [Structural protein] Latent TGF-beta binding protein, may assemble latent TGF-beta complexes in developing elastic tissues, contains proline/glycine-rich sequences alternating with cysteine-rich clusters, expressed in embryonic cartilage perichondrium and blood vessel |
| 5 | 7505057CD1 | g992950 | 5.5E-142 | [Homo sapiens] OPN-c Saitoh, Y. et al. (1995) Expression of osteopontin in human glioma. Its correlation with the malignancy. Lab. Invest. 72: 55-63 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|--|
| 5 cont | | 341294 SPP1 | 1.0E-131 | [Homo sapiens] [Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Secreted phosphoprotein-1 (osteopontin), an extracellular matrix glycoprotein and integrin ligand, plays a role in cellular adhesion and migration, regulates normal and atherosclerotic calcification processes, contributes to tumor invasion and metastasis Gillespie, M. T. et al. (1997) Calcitonin receptors, bone sialoprotein and osteopontin are expressed in primary breast cancers. Int. J. Cancer 73:812-815 |
| | | 584161 Spp1 | 1.3E-69 | [Mus musculus] [Ligand] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Secreted phosphoprotein-1 (osteopontin), an extracellular matrix glycoprotein and integrin ligand, promotes cellular adhesion and migration, inhibits apoptosis, regulates bone resorption, and contributes to the metastatic potential of transformed cells |
| | | 329118 Spp1 | 3.8E-54 | [Rattus norvegicus] [Structural protein] [Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Secreted phosphoprotein-1 (osteopontin), an extracellular matrix glycoprotein and integrin ligand, promotes cellular adhesion and migration, inhibits apoptosis, regulates bone resorption, and contributes to the metastatic potential of transformed cells |
| | | 319604 Dspp | 1.8E-10 | [Mus musculus] [Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Dentin sialophosphoprotein, an extracellular matrix protein of dentin in tooth |
| | | 435702 DSPP | 3.2E-10 | [Homo sapiens] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Dentin sialophosphoprotein (phosphoryn), an extracellular matrix protein of dentin in teeth, contains RGD sequence motifs, may function in dentin mineralization; gene mutation is associated with dentinogenesis imperfecta Shields type II |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 6 | 90116002CD1 | g177179 | 1.1E-58 | [Homo sapiens] alpha-2 type VIII collagen Muragaki, Y. et al. (1991) The alpha 2(VIII) collagen gene. A novel member of the short chain collagen family located on the human chromosome 1. J. Biol. Chem. 266:7721-7727 |
| | | 347468 COL8A2 | 9.3E-60 | [Homo sapiens] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Basement membrane (extracellular matrix); Extracellular (excluding cell wall)] Alpha 2 subunit of type VIII collagen, member of a family of extracellular matrix proteins, expressed in corneal endothelial cells |
| | | 344080 COL10A1 | 9.6E-58 | [Homo sapiens] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Alpha 1 subunit of type X collagen, involved in skeletal development; mutation of the corresponding gene causes Schmid metaphyseal chondrodysplasia |
| | | 322406 Col10a1 | 7.7E-56 | [Mus musculus] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Alpha 1 subunit of type X collagen, involved in skeletal and lymphoid organ development and hematopoiesis; mutation of the corresponding human COL10A1 gene causes Schmid metaphyseal chondrodysplasia |
| | | 429540 Col8a1 | 1.6E-55 | [Mus musculus] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Alpha 1 subunit of type VIII collagen, an extracellular matrix protein |
| | | 334758 COL8A1 | 1.8E-54 | [Homo sapiens] [Structural protein] [Extracellular matrix (cuticle and basement membrane)]; Extracellular (excluding cell wall)] Alpha 1 subunit of type VIII collagen, an extracellular matrix protein that forms homotrimers |
| 7 | 039283CD1 | g15145797 | 5.8E-41 | [Sus scrofa] basic proline-rich protein |
| | | 626566 Ptp | 5.0E-28 | [Mus musculus] [Extracellular (excluding cell wall)] Proline rich protein with tandem repeats, expression is induced in salivary glands by isoproterenol and feeding tannins |
| | | 241434 C50F7.2 | 9.4E-28 | [Caenorhabditis elegans] Protein with similarity to human MUC1 protein, mucin 1, transmembrane protein |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 7 cont | | 341800 PRB1 | 2.7E-24 | [Homo sapiens] [Extracellular (excluding cell wall)] Basic proline-rich salivary protein Kim, H. S. et al. (1993) The structure and evolution of the human salivary proline-rich protein gene family. Mamm. Genome 4:3-14 |
| | | 326772 Mm.30280 | 2.7E-23 | [Mus musculus] Proline-rich protein (MP2), an acidic proline-rich protein which is a member of a family of proline-rich proteins expressed in salivary gland, induced by isoproterenol and also regulated by isoprenaline |
| 8 | 7505082CD1 | g5456977 | 1.8E-96 | [Homo sapiens] protocadherin gamma C3. Wu, Q. and Maniatis, T. (1999) Cell 97:779-790 |
| | 7505082CD1 | 336796 PCDH2 | 4.6E-81 | [Homo sapiens][Adhesin/agglutinin][Plasma membrane; Cell junction] Protocadherin 2, a cadherin-related molecule that is expressed in the central nervous system and localizes to cell-cell contact sites, acts as an adhesin. Matsuyoshi, N. and Imamura, S. (1997) Biochem. Biophys. Res. Commun. 235:355-358. |
| 9 | 7505139CD1 | g3941728 | 1.8E-81 | [Homo sapiens] sialomucin CD164. Watt, S.M. et al. (1998) Blood 92: 849-866 |
| | 7505139CD1 | 342954 CD164 | 4.1E-66 | [Homo sapiens][Adhesin/agglutinin][Lysosome/vacuole; Endosome/Endosomal vesicles; Cytoplasmic; Plasma membrane] Sialomucin, a cell surface glycoprotein that plays a role in hematopoietic progenitor cell adhesion and proliferation, ratio of soluble and transmembrane variant forms correlates with metastatic potential in colorectal carcinomas Matsui, T. et al. (2000) J. Biochem. (Tokyo) 127:1103-7 |
| 10 | 7505234CD1 | g5457039 | 5.3E-253 | [Homo sapiens] protocadherin beta 2 Wu, Q. and Maniatis, T. (1999) <i>supra</i> |
| | 7505234CD1 | 605860 PCDHB2 | 4.6E-254 | [Homo sapiens][Adhesin/agglutinin][Plasma membrane] Member of a family of neural cadherin-like cell adhesion proteins, may have a role in forming neuronal connections in the brain; corresponding gene is one of a cluster of genes encoding related cadherin-like proteins Wu, Q., and Maniatis, T. (2000) Proc. Natl. Acad. Sci. U. S. A. 97:3124-3129 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 11 | 7500227CD1 | g16118489 | 3.2E-11 | [Gallus gallus] slit-2 Holmes, G. and Niswander, L. (2001) Dev. Dyn. 222:301-307 |
| | 7500227CD1 | 325750 Tpbg | 3.5E-10 | [Mus musculus][Plasma membrane] Trophoblast glycoprotein, likely involved in cell adhesion and motility and pregnancy; human 5T4 is strongly associated with tumor metastasis King, K. W. et al. (1999) Biochim. Biophys. Acta 1445:257-270 |
| 12 | 7503676CD1 | g2865219 | 4.2E-32 | [Homo sapiens] integrin binding protein Del-1 Hidai, C. et al. (1998) Genes Dev. 12:21-33 |
| | 7503676CD1 | 342430 EDIL3 | 3.7E-33 | [Homo sapiens][Ligand][Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] EGF-like repeats and discoidin I-like domains 3, may be a ligand for the alpha 5 beta 3 integrin receptor, may promote endothelial cell attachment and migration, may be involved in vascular morphogenesis or remodeling. Hidai, C. et al. (1998) supra |
| 13 | 7503606CD1 | g600118 | 2.0E-34 | [Zea mays] extensin-like protein |
| | | 424290 K1A0229 | 7.6E-41 | [Homo sapiens] Protein containing six ankyrin (Ank) repeats and two SAM (sterile alpha motif) domains, which may mediate protein-protein interactions, contains a phosphotyrosine interaction domain (PTB/PID) |
| | | 624398 Prosap2 | 5.3E-26 | [Rattus norvegicus][Ligand][Cytoskeletal] Proline-rich synapse-associated protein-1 (cortactin binding protein 1), a protein with a PDZ-domain that may function in postsynaptic density assembly during neuronal differentiation |
| 14 | 7500216CD1 | g10334770 | 0.0 | [Homo sapiens] MUCDHL-FL Paris, M.J. and Williams, B.R.G. (2000) Genomics 69:196-202 |
| | | 629545 Rn.18860 | 1.5E-228 | [Rattus norvegicus][Receptor (signalling)][Basolateral plasma membrane] mu-Protocadherin, a cell-cell adhesion molecule that may participate in regulation of branching morphogenesis during organ development Goldberg, M. et al. (2000) J. Biol. Chem. 275:24622-24629 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|--|
| 14 cont | | 731989 PCDH15 | 2.5E-16 | [Homo sapiens] Protocadherin 15, involved in the maintenance of cochlear and retinal functions; mutations in the corresponding gene cause Usher syndrome type 1F Alagramam, K. N. et al. (2001) Hum. Mol. Genet. 10:1709-1718 |
| 15 | 7099880CD1 | g9845485 | 0.0 | [Homo sapiens] protocadherin-9 |
| | | 610944 PCDH9 | 0.0 | [Homo sapiens][Ligand; Receptor (signalling)][Plasma membrane] Protein with high similarity to PCDH11 (protocadherin 11), which is a cadherin-related molecule involved in cell-cell recognition in the central nervous system; contains six cadherin domains found in the some receptors and cell adhesion proteins |
| | 7099880CD1 | 743170 PCDH11 | 0.0 | [Homo sapiens][Adhesin/agglutinin][Plasma membrane] Protocadherin 11, a cadherin-related molecule involved in cell-cell recognition in the central nervous system |
| 16 | 871513CD1 | g6636372 | 5.0E-154 | [Caenorhabditis elegans] UNC-112 |
| | | 424562 MIG2 | 0.0 | Rogalski, T.M. et al. (2000) J. Cell Biol. 150:253-264 |
| | | 430352 Tln | 1.4E-16 | [Homo sapiens] Protein whose expression is induced by mitogen |
| | | | | [Mus musculus][Anchor Protein][Cytoplasmic; Plasma membrane; Cytoskeletal; Cell junction] Talin, a component of adherens junctions that links the actin cytoskeleton to the cytoplasmic domain of integrin receptors and may function in cell adhesion |
| 17 | 8057640CD1 | g33957 | 0.0 | Martel, V. et al. (2001) J. Biol. Chem. 276:21217-2127 |
| | | 331028 Itgb4 | 0.0 | [Homo sapiens] integrin beta 4 subunit |
| | | | | Hogervorst, F. et al. (1990) Cloning and sequence analysis of beta-4 cDNA: an integrin subunit that contains a unique 118 kd cytoplasmic domain. EMBO J. 9:765-770 |
| | | | | [Rattus norvegicus][Adhesin/agglutinin; Receptor (signalling)][Plasma membrane] Integrin beta 4 subunit, binds integrin alpha 6 (ITGA6) to form a receptor that anchors epithelial cells to the basal lamina, decreased expression contributes to metastasis of carcinomas; deficiency of human ITGB4 is associated with epidermolysis bullosa |
| | | | | Khare, L. et al. (1998) Alterations in the expression of alpha6beta4 integrin and p21/WAF1/Cip1 in N-nitrosomethylbenzylamine-induced rat esophageal tumorigenesis. Mol. Carcinog. 21:185-193 |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 17 cont | | 336064 ITGB4 | 0.0 | [Homo sapiens][Adhesin/agglutinin;Receptor (signalling)][Plasma membrane] Integrin beta 4 subunit, binds integrin alpha 6 (ITGA6) to form a receptor that anchors epithelial cells to the basal lamina, loss of function contributes to metastasis of carcinomas, deficiency is associated with epidermolysis bullosa and pyloric atresia Niessen, C. M. et al. (1996) Deficiency of the integrin beta 4 subunit in junctional epidermolysis bullosa with pyloric atresia: consequences for hemidesmosome formation and adhesion properties. [published erratum appears in J. Cell Sci. (1996) 109(Pt 8):preceding table of contents] J. Cell Sci.:1695-1706 |
| 18 | 7505913CD1 | g1575766 | 1.30E-200 | [Homo sapiens] cytohesin-2 Frank, S. et al. (1998) ARNO is a guanine nucleotide exchange factor for ADP-ribosylation factor 6. J. Biol. Chem. 273:23-27 |
| | | 571452 PSCD2 | 1.10E-201 | [Homo sapiens][Guanine nucleotide exchange factor][Plasma membrane] Guanine nucleotide exchange factor for ARF6, which has roles in endocytosis and in assembly of the actin cytoskeleton, localized to the plasma membrane, also has activity against ARF1 Mukherjee, S. et al. (2000) The ADP ribosylation factor nucleotide exchange factor ARNO promotes beta -arrestin release necessary for luteinizing hormone/choriogonadotropin receptor desensitization. Proc. Natl. Acad. Sci. U S A 97:5901-5906 |
| | | 585609 Pscd2 | 1.60E-200 | [Mus musculus][Guanine nucleotide exchange factor][Plasma membrane] Protein with very strong similarity to human PSCD2, which is a guanine nucleotide exchange factor for ARF6 and that localizes to the plasma membrane Kim, H. S. et al. (1998) Complex regulation of multiple cytohesin-like genes in murine tissues and cells. FEBS Lett. 433:312-316 |
| 19 | 7510292CD1 | g11559520 | 2.0E-79 | [Drosophila melanogaster] extracellular matrix protein papilin Kramerova, I. A. et al. Development 127, 5475-5485 (2000) |
| 20 | 7504669CD1 | g5764665 | 3.3E-178 | [Homo sapiens] cerebral cell adhesion molecule Starzyk, R. et al. J. Infect. Dis. 181, 181-187 (2000) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 20 cont | | 475793 LOC511428 | 2.8E-179 | [Homo sapiens][Adhesin/agglutinin][Plasma membrane; Unspecified membrane] Cerebral cell adhesion molecule, an adhesion molecule that is predicted to be involved in leukocyte transmigration across the blood-brain barrier |
| | | 430154 P1od3 | 4.5E-13 | [Mus musculus][Oxidoreductase] Lysine hydroxylase, catalyzes the formation of hydroxylysine in collagens |
| | | 337038 P1od3 | 2.6E-12 | [Homo sapiens][Oxidoreductase] Lysine hydroxylase, catalyzes the formation of hydroxylysine in collagens |
| 21 | 7509266CD1 | g9446402 | 2.0E-28 | [Homo sapiens] integrin beta-subunit |
| | | 606202 ITGB6 | 1.7E-29 | [Homo sapiens][Adhesin/agglutinin; Receptor (signalling)][Plasma membrane] Integrin beta 6, member of a family of cell-surface proteins, binds fibronectin, mediates epithelial cell-matrix interactions in development, wound repair, and neoplasia, regulates lung inflammatory response, receptor for foot and mouth disease virus |
| 22 | 7509288CD1 | g9446402 | 0.0 | [Homo sapiens] integrin beta-subunit |
| | | | | Sheppard, D. et al. <u>supra</u> |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|---------------------------|--------------------------|---|----------------------|---|
| 22 cont | | 606202 TGB6 | 0.0 | [Homo sapiens][Adhesin/agglutinin; Receptor (signalling)][Plasma membrane] Integrin beta 6, member of a family of cell-surface proteins, binds fibronectin, mediates epithelial cell-matrix interactions in development, wound repair, and neoplasia, regulates lung inflammatory response, receptor for foot and mouth disease virus Krissansen, G. W. et al. (supra) Sheppard, D. et al. (supra) Boukerche, H. et al. (supra) Jackson, T. et al. (supra) Agrez, M. et al. (supra) Agrez, M. et al. (supra) Weinacker, A. et al. (supra) Berthet, V. et al. (supra) Koivisto, L. et al. (supra) |
| | | 339440 TGB3 | 2.8E-204 | [Homo sapiens][Adhesin/agglutinin; Isomerase; Chaperones; Receptor (signalling); Small molecule-binding protein][Plasma membrane; Cell junction] Integrin beta 3 (glycoprotein IIIa), subunit of integrin, involved in platelet aggregation, cell adhesion, blood coagulation, and cell proliferation and differentiation; gene mutation is associated with the bleeding disorder Glanzmann thrombasthenia Zimrin, A. B. et al. J Clin Invest 81, 1470-5 (1988) Bray, P. F. et al. Proc Natl Acad Sci U S A 85, 8683-7 (1988) Chen, Y. P. et al. Proc Natl Acad Sci U S A 89, 10169-73 (1992) Longhurst, C. M. et al. Eur J Biochem 263, 104-11 (1999) Savage, B. et al. Cell 94, 657-66 (1998) Furman, M. I. et al. Proc Natl Acad Sci U S A 95, 3082-7 (1998) Polgar, J. et al. J Biol Chem 272, 13576-83 (1997) Zhang, J. et al. J Biol Chem 271, 6265-72 (1996) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|---------------------------|--------------------------|---|----------------------|---|
| 22 cont | | 587959 Itgb3 | 4.6E-202 | [Mus musculus][Adhesin/agglutinin; Receptor (signalling)][Plasma membrane] Integrin beta 3, subunit of integrin, involved in platelet aggregation, cell adhesion, blood coagulation, and likely in cell proliferation and differentiation; mutations in human ITGB3 cause the bleeding disorder Glanzmann thrombasthenia Ramakrishnan, V. et al. Proc Natl Acad Sci U S A 96, 13336-41 (1999) Silletti, S. et al. J Cell Biol 149, 1485-502 (2000) Chen, N. et al. J Biol Chem 275, 24953-61 (2000) Brakebusch, C. et al. Embo Journal 19, 3990-4003 (2000) Ohlmann, P. et al. Blood 96, 2134-9 (2000) Yun, Z. et al. Cancer Res 56, 3103-11 (1996) Nakamura, I. et al. Endocrinology 139, 5182-93 (1998) Illera, M. J. et al. Biol Reprod 62, 1285-90 (2000) |
| 23 | 7510212CD1 | g1621019 | 5.8E-279 | [Homo sapiens] fibulin-1D Tran, H. et al. Matrix Biol. 15, 479-493 (1997) |
| | | 568154 FBLN1 | 4.7E-280 | [Homo sapiens][Structural protein][Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Fibulin 1, an extracellular matrix and plasma glycoprotein that may connect elements of the extracellular matrix, may play a role in hemostasis, altered expression may play a role in tumor invasion, thrombosis, and connective tissue and blood diseases Pan, T. C. et al. Eur J Biochem 215, 733-740 (1993) Talts, J. F. et al. J Cell Sci, 2153-2162 (1995) Argraves, W. S. et al. J Cell Biol 111, 3155-3164 (1990) Korenberg, J. R. et al. Cytogenet Cell Genet 68, 192-193 (1995) Tran, H. et al. J Biol Chem 270, 19458-19464 (1995) Qing, J. et al. Oncogene 15, 2159-2168. (1997) Hayashido, Y. et al. Int J Cancer 75, 654-658. (1998) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 23 cont | | 584763 Fbln1 | 9.3E-236 | [Mus musculus][Structural protein][Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Fibulin 1, an extracellular matrix glycoprotein that may connect elements of the extracellular matrix, may play roles in heart, lung, and kidney development, epithelial mesenchymal transitions, tensile strength of cardiac valves, and tumor invasion Pan, T. C. et al. J Cell Biol 123, 1269-77 (1993) Pan, T. C. et al. FEBS Lett 444, 38-42 (1999) Zhang, H. Y. et al. Dev Biol 167, 18-26 (1995) Zhang, H. Y. et al. Dev Dyn 205, 348-64 (1996) Fassier, R. et al. Exp Cell Res 222, 111-6 (1996) Talts, J. F. et al. J Biol Chem 275, 35192-9 (2000) Olin, A. I. et al. J Biol Chem 276, 1253-1261. (2001) Miosge, N. et al. FASEB J 13, 1743-50. (1999) [Caenorhabditis elegans] Member of the EGF-repeat protein family Barth, J. L. et al. Matrix Biol 17, 635-46 (1998) Walhout, A. J. et al. Science 287, 116-22 (2000) |
| 24 | 7510504CD1 | g15099921 | 0.0 | [Homo sapiens] ADAM-TS related protein 1 (A Disintegrin-like And Metalloproteinase domain with Thrombospondin type I modules Related gene-1) |
| | | 610229 KIAA1233 | 1.8E-171 | [Homo sapiens] Protein containing thirteen ankyrin (Ank) repeats, which may mediate protein-protein interactions |
| | | 243710 gon-1 | 8.5E-99 | [Caenorhabditis elegans][Protease (other than proteasomal)][Extracellular (excluding cell wall)] Metalloprotease required for gonad morphogenesis Abbaszade, I. et al. J Biol Chem 274, 23443-50 (1999) Blelloch, R. et al. Nature 399, 586-590 (1999) Blelloch, R. et al. Dev Biol 216, 382-93 (1999) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|--|
| 24 cont | | 703727 ADAMT S12 | 2.4E-85 | [Homo sapiens] Member of the disintegrin-like and metalloprotease (repolysin type) with thrombospondin type 1 motif family of proteases, expressed in fetal lung Cal. S. et al. J Biol Chem 276. 17932-40. (2001) |
| 25 | 7510587 CD1 | g6009533 | 3.1E-241 | [Homo sapiens] tubulointerstitial nephritis antigen Ikeda, M. et al. Biochem. Biophys. Res. Commun. 268, 225-230 (2000) |
| | | 569510 TINAG | 2.5E-242 | [Homo sapiens][Hydrolase; Protease (other than proteasomal)][Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix)] Tubulointerstitial nephritis antigen, kidney-specific basement membrane protein that contains an ATP/GTP binding site and cysteine residues within a follistatin module, functions in renal tubulogenesis; defective in hereditary tubulointerstitial disorders Ikeda, M. et al. Biochem Biophys Res Commun 268, 225-30 (2000) Wex, T. et al. Biochemistry 40, 1350-7. (2001) |
| | | 430346 Tinag | 7.1E-206 | [Mus musculus][Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix)] Tubulointerstitial nephritis antigen, basement membrane protein of renal tubules that contains an ATP/GTP binding site and selectively regulates renal tubulogenesis; human TINAG is defective in hereditary tubulointerstitial disorders Kanwar, Y. S. et al. Proc Natl Acad Sci U S A 96, 11323-8 (1999) |
| | | 243789 F26E4.3 | 7.7E-74 | [Caenorhabditis elegans] Member of the cysteine protease protein family |
| 26 | 7510684 CD1 | g6009533 | 2.1E-224 | [Homo sapiens] tubulointerstitial nephritis antigen Ikeda, M. et al. (supra) |
| | | 569510 TINAG | 1.7E-225 | [Homo sapiens][Hydrolase; Protease (other than proteasomal)][Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix)] Tubulointerstitial nephritis antigen, kidney-specific basement membrane protein that contains an ATP/GTP binding site and cysteine residues within a follistatin module, functions in renal tubulogenesis; defective in hereditary tubulointerstitial disorders Ikeda, M. et al. (supra) Wex, T. et al. (supra) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 26 cont | | 430346 Tinag | 1.6E-200 | [Mus musculus][Extracellular matrix (cuticle and basement membrane); Basement membrane (extracellular matrix)] Tubulointerstitial nephritis antigen, basement membrane protein of renal tubules that contains an ATP/GTP binding site and selectively regulates renal tubulogenesis; human TINAG is defective in hereditary tubulointerstitial disorders Kanwar, Y. S. et al. (supra) |
| | | 243789 F26E4.3 | 6.5E-77 | [Caenorhabditis elegans] Member of the cysteine protease protein family |
| 27 | 7510697CD1 | g6164595 | 7.0E-93 | [Manduca sexta] lacunin |
| | | 691408 FLJ13710 | 2.6E-121 | Nardi, J. B. et al. Insect Biochem. Mol. Biol. 29, 883-897 (1999) |
| | | | | [Homo sapiens] Protein containing six type 1 thrombospondin domains, has a region of low similarity to a region of A disintegrin-like and metalloprotease 2 (human ADAMTS2), which is a collagenase associated with Ehlers-Danlos syndrome type VIIC upon gene mutation |
| 28 | 7761337CD1 | g556845 | 0.0 | [Homo sapiens] human tenascin-C |
| | | | | Gherzi, R. et al. J. Biol. Chem. 270, 3429-3434 (1995) |
| 29 | 7503666CD1 | g10334774 | 0.0 | [Homo sapiens] MUCDHL-FL |
| | 7503666CD1 | 659060 MUCDH L | 0.0 | [Homo sapiens] Mucin and cadherin like, protein of unknown function, has a region of low similarity to proline-rich repeat proteins; corresponding gene is in region associated with many cancers |
| | | | | Paris, M. J., and Williams, B. R. Genomics 69:196-202 (2000) |
| | 7503666CD1 | 629545 Rn.18860 | 3.2E-249 | [Rattus norvegicus][Receptor (signalling)][Basolateral plasma membrane] mu-Protocadherin, a cell-cell adhesion molecule that may participate in regulation of branching morphogenesis during organ development |
| | | | | Goldberg, M., et al. J. Biol. Chem. 275:24622-9 (2000) |
| 30 | 7503668CD1 | g10334774 | 0.0 | [Homo sapiens] MUCDHL-FL |
| | 7503668CD1 | 659060 MUCDH L | 0.0 | [Homo sapiens] Mucin and cadherin like, protein of unknown function, has a region of low similarity to proline-rich repeat proteins; corresponding gene is in region associated with many cancers |
| | | | | Paris, M. J., and Williams, B. R. (2000) <u>supra</u> |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 30 cont | 7503668CD1 | 629545[Rn.18860] | 4.9E-261 | [Rattus norvegicus][Receptor (signalling)][Basolateral plasma membrane] mu-Protocadherin, a cell-cell adhesion molecule that may participate in regulation of branching morphogenesis during organ development Goldberg, M., et al. (2000) <i>supra</i> |
| 31 | 7503672CD1 | g10334774 | 0.0 | [Homo sapiens] MUCDHL-FL |
| | 7503672CD1 | 659060[MUCDHL] | 0.0 | [Homo sapiens] Mucin and cadherin like, protein of unknown function, has a region of low similarity to proline-rich repeat proteins; corresponding gene is in region associated with many cancers Paris, M. J., and Williams, B. R. (2000) <i>supra</i> |
| | 7503672CD1 | 629545[Rn.18860] | 3.1E-212 | [Rattus norvegicus][Receptor (signalling)][Basolateral plasma membrane] mu-Protocadherin, a cell-cell adhesion molecule that may participate in regulation of branching morphogenesis during organ development Goldberg, M., et al. (2000) <i>supra</i> |
| 32 | 6039650CD1 | g930343 | 0.0 | [Homo Sapiens] LAR-interacting protein 1b Serra-Pages, C. et al. EMBO J. 14, 2827-2838 (1995) |
| | 6039650CD1 | 337114[PPF1A1] | 0.0 | [Homo sapiens][Cytoplasmic; Plasma membrane; Cell junction] Protein tyrosine phosphatase receptor type f polypeptide (PTPRF) interacting protein (liprin) alpha 1, localizes the PTPRF transmembrane protein tyrosine phosphatase to focal adhesions; also interacts with PTPRS and PTPRD Serra-Pages, C. et al. <i>supra</i> ; Pulido, R. et al. Proc. Natl. Acad. Sci. USA 92, 11686-90 (1995) |
| | 6039650CD1 | 337116[PPF1A2] | 0.0 | [Homo sapiens][Cytoplasmic; Plasma membrane; Cell junction] Liprin alpha 2, binds and localizes LAR family protein tyrosine phosphatase members LAR (PTPRF), PTP alpha (PTPRA), and PTP delta (PTPRD) at focal adhesions, plays a role in the regulation of cell-matrix interactions Serra-Pages, C. et al. J. Biol. Chem. 273, 15611-20 (1998) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 33 | 7509919CD1 | g825636 | 1.2e-21 | [Homo sapiens] integrin beta-2 subunit Weitzman, J. B. et al. FEBS Lett. 294, 97-103 (1991) |
| | 7509919CD1 | 339642 ITGB2 | 9.8e-23 | [Homo sapiens][Adhesin/agglutinin; Receptor (signaling); Small molecule-binding protein][Plasma membrane] Beta 2 subunit of integrin; member of a family of cell-surface proteins involved in cell-cell and cell-matrix interactions; mutation of the corresponding gene causes leukocyte adhesion deficiency Back, A. L. et al. Biochem. Biophys. Res. Comm. 193, 912-8 (1993); Weitzman, J. B. et al. supra; Wardlaw, A. J. et al. J. Exp. Med. 172, 335-45 (1990); Bella, J. et al. Proc. Natl. Acad. Sci. USA 95, 4140-5 (1998) |
| | 7509919CD1 | 583431 Itgb2 | 3.40E-16 | [Mus musculus][Receptor (signaling)][Plasma membrane] Beta 2 subunit of integrin, a component of macrophage activation antigen 1 (complement receptor type 3) and member of a family of cell-surface proteins involved in cell-cell and cell-matrix interactions Zeger, D. L. et al. Immunogenetics 31, 191-7 (1990); Puijt, J. F. et al. Blood 93, 107-12 (1999); Imai, Y. et al. Immunology 86, 591-8 (1995) |
| 34 | 7510758CD1 | g14575679 | 7.00E-31 | [Homo sapiens] hemiscentin |
| | 7510758CD1 | 610613 LOC57164 | 9.80E-90 | [Homo sapiens] Protein containing a MAM domain, which are adhesive domains found in many receptor families |
| | 7510758CD1 | 329980 Rn.10117 | 4.90E-32 | [Rattus norvegicus][Adhesin/agglutinin] Axonal-associated cell adhesion molecule, member of TAG 1/F3 subgroup of the immunoglobulin superfamily, a cell adhesion molecule that promotes neurite outgrowth and may play a role in establishing neuronal networks in the brain Yoshihara, Y. et al. J. Neurobiol. 28, 51-69 (1995) |
| 35 | 7510063CD1 | g632776 | 0.0 | [Homo sapiens] N-CAM Saito, S. et al. Lung Cancer 10, 307-318 (1994) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|------------------------|-----------------------|-----------------------------------|-------------------|---|
| 35 cont | 7510063CD1 | 618388[NCAM1] | 0.0 | [Homo sapiens][Adhesin/agglutinin][Plasma membrane] Neural cell adhesion molecule 1, a member of the immunoglobulin superfamily, important for cell adhesion between neurons and at neuromuscular junctions, functions to regulate axonal outgrowth and neuronal remodeling Rutishauser, U. et al. Science 240, 53-7 (1988); Cunningham, B. A. et al. Science 236, 799-806 (1987) |
| | 7510063CD1 | 711370[Ncam1] | 0.0 | [Rattus norvegicus][Adhesin/agglutinin][Plasma membrane] Neural cell adhesion molecule, a member of the immunoglobulin superfamily, important for homophilic cell adhesion between neurons and at neuromuscular junctions, functions to regulate axonal outgrowth and neuronal remodeling Miura, M. et al. FEBS Lett. 289, 91-5 (1991); Small, S. J. et al. J. Cell Biol. 105, 2335-45 (1987) |
| 36 | 7510135CD1 | g340307 | 2.5e-106 | [Homo sapiens] vitronectin alpha subunit precursor Suzuki, S. et al. J. Biol. Chem. 262, 14080-14085 (1987) |
| | 7510135CD1 | 336058[ITGAV] | 2.0e-107 | [Homo sapiens][Adhesin/agglutinin; Isomerase; Chaperones; Receptor (signaling); Small molecule-binding protein][Plasma membrane; Cell junction] Alpha V subunit integrin, a subunit of the vitronectin receptor that is involved in cell-cell and cell-matrix interactions; plays a role in tumor angiogenesis and may contribute to tumorigenicity of cutaneous malignant melanoma Maeshima, Y. et al. J. Biol. Chem. 275, 23745-50 (2000); Sims, M. A. et al. Cytogenet. Cell Genet. 89, 268-71 (2000) |
| | 7510135CD1 | 583427[Itgav] | 3.2e-93 | [Mus musculus][Receptor (signaling)][Plasma membrane] Alpha V subunit integrin, a subunit of the vitronectin receptor that is involved in cell-cell and cell-matrix interactions; human ITGAV may play a role in tumorigenicity of cutaneous malignant melanoma Wada, J. et al. J. Cell Biol. 132, 1161-76 (1996); Bader, B. L. et al. Cell 95, 507-19 (1998) |
| 37 | 7505011CD1 | g556845 | 0.0 | [Homo sapiens] human tenascin-C Gherzi, R. et al. J. Biol. Chem. 270, 3429-3434 (1995) |

Table 2

| Polypeptide SEQ ID NO: | Incyte Polypeptide ID | GenBank ID NO: or PROTEOME ID NO: | Probability Score | Annotation |
|---------------------------|--------------------------|---|----------------------|--|
| 37 cont | 7505011CD1 | 335910 HXB | 0.0 | [Homo sapiens][Adhesin/agglutinin; Receptor (signaling)][Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Tenascin C (hexabrachion), a multidomain extracellular matrix glycoprotein, may be involved in cell adhesion and neurite outgrowth, induced by hypoxia and has increased expression in colorectal carcinoma, osteosarcoma metastases and pulmonary hypertension Nies, D. E. et al. J. Biol. Chem. 266, 2818-23 (1991); Gulcher, J. R. et al. Proc. Natl. Acad. Sci. USA 88, 9438-42 (1991); Wenk, M. B. et al. J. Cell Biol. 150, 913-20 (2000) |
| | 7505011CD1 | 586659 Tnc | 0.0 | [Mus musculus][Structural protein][Extracellular matrix (cuticle and basement membrane); Extracellular (excluding cell wall)] Tenascin C (hexabrachion), a multidomain extracellular matrix glycoprotein with epidermal growth factor-like and fibronectin type III-like repeats, may be involved in cell adhesion. has effects on cell proliferation and migration Weller, A. et al. J. Cell Biol. 112, 355-62 (1991); Fukamauchi, F. et al. Neurochem. Int. 36, 153-8 (2000); Sakai, T. et al. J. Cell Sci. 2069-77 (1996) |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 1 | 7504868CD1 | 311 | S121 S216 S239 S287 T45 T91 T112 T133 T141 T149 T157 T165 T173 T181 T189 T197 T210 T219 T230 | N147 | signal_cleavage: M1-A18 | SPSCAN |
| | | | | | Signal Peptide: M1-A18, M1-P20, M1-Q22, M1-V25 | HMMER |
| | | | | | CELL ADHESION ADDRESSIN MADCAM-1 MOLECULE MUCOSAL MOLECULE 1 SMALL FORM VASCULAR PD009780: L15-I126 | BLAST_PRODOM |
| 5 | | | | | ADDRESSIN VASCULAR MADCAM-1 MUCOSAL CELL ADHESION MOLECULE 1 PD029871: A243-S311 | BLAST_PRODOM |
| | | | | | H-A-P-P REPEAT (Extensin precursor (Cell wall hydroxyproline-rich glycoprotein) in tobacco & tomato) DM08271 S25299 69-249: T91-P214 | BLAST_DOMO |
| | | | | | Leucine zipper pattern: L9-L30 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---|--|----------------------------------|
| 2 | 7504930CD1 | 1445 | S46 S64 S81 S156 S294 S451 S606 S620 S677 S731 S923 S1041 S1082 S1132 S1198 S1276 S1278 S1286 S1322 S1323 S1380 S1402 S1418 T143 T212 T279 T311 T365 T371 T458 T532 T581 T603 T628 T759 T784 T808 T831 T853 T857 T876 T908 T932 T1035 T1112 T1116 T1182 T1409 Y127 Y408 Y874 | N73 N210 N326 N470 N489 N639 N715 N893 N1130 N1282 | signal_cleavage: M1-A33 | SPSCAN |
| | | | | | Signal Peptide: M1-G30, M1-A33 | HMMER |
| | | | | | Fibronectin type III domain: P539-T621, P633-T721, P938-S1028, P439-S525, P739-L821, P837-S926 | HMMER_PFAM |
| | | | | | Immunoglobulin domain: G263-A322, G166-V223, G67-A131, S355-A412 | HMMER_PFAM |
| | | | | | Cytosolic domain: T1112-A1445 Transmembrane domain: L1089-C1111 Non-cytosolic domain: M1-M1088 | TMHMMER |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 2 | | | | | Receptor tyrosine kinase class V proteins BL00790: V450-F476, Y477-G520, S554-K579 | BLIMPS_BLOCKS |
| cont | | | | | Fibronectin type III repeat signature PR00014: T752-P761, A892-Y910, Y1012-P1026 | BLIMPS_PRINTS |
| | | | | | TUMOR SUPPRESSOR NEOGENIN PROTEIN DCC PRECURSOR COLORECTAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD PD041287: D1164-T1443 | BLAST_PRODROM |
| | | | | | NEOGENIN PROTEIN PD024613: E1029-L1090 | BLAST_PRODROM |
| | | | | | TUMOR SUPPRESSOR NEOGENIN PROTEIN DCC PRECURSOR COLORECTAL GLYCOPROTEIN IMMUNOGLOBULIN FOLD PD009999: C1111-P1167 | BLAST_PRODROM |
| | | | | | NEOGENIN PROTEIN PD020198: M1-R66 | BLAST_PRODROM |
| | | | | | IMMUNOGLOBULIN DM00001 | BLAST_DOMO |
| | | | | | P43146 328-410: P341-Q420 | |
| | | | | | P43146 42-127: F55-I140 | |
| | | | | | FIBRONECTIN TYPE III REPEAT DM00007 | BLAST_DOMO |
| | | | | | P43146 935-1014: A929-D1009 | |
| | | | | | P43146 834-912: P829-N907 | |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|--|---|----------------------------------|
| 3 | 6610456CD1 | 726 | S73 S224 S243 S260 S285 S297 S317 S325 S332 S401 S416 S420 S438 S442 S443 S448 S450 S461 S491 S496 S500 S508 S513 S539 S546 S616 S618 S659 S706 T82 T90 T126 T156 T166 T233 T331 T358 T362 T524 T707 Y378 | N7 N48 N117 N259 N337 N436 N520 N590 N603 N704 | signal_cleavage: M1-G23 | SPSCAN |
| | | | | | Signal Peptide: M1-G23 | HMMER |
| | | | | | 5'-nucleotidase, catalytic domain: E98-N117 | HMMER_PFAM |
| | | | | | Guanylate-binding protein, C-terminal: H86-K114 | HMMER_PFAM |
| | | | | | Zinc finger, C3HC4 type (RING finger): C18-C55 | HMMER_PFAM |
| | | | | | Leucine zipper pattern: L181-L202 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 4 | 7503573CD1 | 1474 | S63 S291 S297 S301 S317 S354 S431 S572 S631 S656 S668 S753 S820 S833 S880 S888 S921 S938 S947 S991 S1084 S1182 S1193 S1358 S1374 S1449 T77 T309 T460 T470 | N315 N388 N1018 N1189 | signal_cleavage: M1-R17 | SPSCAN |
| | | | T498 T796 T815 T899 T965 T1092 T1100 T1139 T1221 T1355 T1406 Y958 | | Signal Peptide: M1-A15 | HMMER |
| | | | | | EGF-like domain: C1150-C1186, C512-C548, C1016-C1052, C324-C359, C719-C755, C596-C632, C887-C922, C845-C881, C116-C143, C638-C670, C677-C713, C1387-C1422, C1428-C1467, C801-P838, C1107-C1144, C1058-C1095, C554-C590, C928-C966, C972-C1010, C761-C795 | HMMER_PFAM |
| | | | | | TB domain (Also known as the 8 cysteine domain. This family includes the hybrid domains. This cysteine rich repeat is found in TGF binding protein and fibrillin.): L1209-L1251, G379-I421, G259-A302 | HMMER_PFAM |

Table 3

| SEQ ID NO: | Incye Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 4 | | | | | EGF-like domain proteins BL00022: C120-C126, C344-F350 | BLIMPS_BLOCKS |
| | | | | | LATENT BINDING GLYCOPROTEIN EGF-LIKE DOMAIN TGF-BETA PROTEIN 4 TRANSFORMING GROWTH FACTOR-BETA PD091047: C1252-G1392 | BLAST_PRODROM |
| | | | | | LATENT BINDING GLYCOPROTEIN EGF-LIKE DOMAIN TGF-BETA PROTEIN 4 TRANSFORMING GROWTH FACTOR-BETA PD067838: C422-M511 | BLAST_PRODROM |
| | | | | | LATENT TGF-BETA BINDING PROTEIN 4 GLYCOPROTEIN EGF-LIKE DOMAIN PD097075: S24-Q105 | BLAST_PRODROM |
| | | | | | LATENT BINDING GLYCOPROTEIN EGF-LIKE DOMAIN TGF-BETA TRANSFORMING GROWTH PROTEIN PROTEIN 2 PD019688: P229-E323 | BLAST_PRODROM |
| | | | | | TGFBP REPEAT DM00210 Q00918 1506-1591: V1200-G1267 P22064 1188-1273: V1200-V1254 | BLAST_DOMO |
| | | | | | EGF-LIKE DOMAIN DM00864 155476 159-241: C608-E680, S895-E974, A982-E1059, C566-Q640, C689-E762 | BLAST_DOMO |
| | | | | | EGF DM00003 Q00918 1609-1652: E1380-D1424 | BLAST_DOMO |
| | | | | | Aspartic acid and asparagine hydroxylation site: C335-C346, C566-C577, C608-C619, C650-C661, C689-C700, C731-C742, C814-C825, C857-C868, C898-C909, C1028-C1039, C1070-C1081, C1120-C1131, C1162-C1173, C1443-C1454 | BLAST_MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|--|----------------------------------|
| 4 | | | | | EGF-like domain signature 1: C132-C143, C659-C670 | MOTIFS |
| cont | | | | | EGF-like domain signature 2: C344-C359, C533-C548, C575-C590, C617-C632, C659-C670, C698-C713, C740-C755, C866-C881, C1037-C1052, C1129-C1144, C1407-C1422, C1452-C1467 | MOTIFS |
| | | | | | Calcium-binding EGF-like domain pattern signature: D320-C344, D550-C575, D592-C617, D634-C659, D673-C698, D715-C740, D757-C781, D797-C823, D841-C866, D883-C907, D924-C950, D968-C994, D1012-C1037, D1054-C1079, D1103-C1129, D1146-C1171, D1424-C1452 | MOTIFS |
| 5 | 7505057CD1 | 273 | S26 S130 S150 S174 S187 S198 S239 S250 T144 Y140 | N38 N65 | signal_cleavage: M1-A16 | SPSCAN |
| | | | | | Signal Peptide: M1-A16, M1-P18, M1-Q21, M1-A22 | HMMER |
| | | | | | Osteopontin: M1-N273 | HMMER_PFAM |
| | | | | | Osteopontin proteins BL00884: M1-K30, F105-E146, S234-S250, E252-N273 | BLIMPS_BLOCKS |
| | | | | | Osteopontin signature PR00216: R2-Q31, T106-D120, I138-E156, S217-D242, E252-E271 | BLIMPS_PRINTS |
| | | | | | PRECURSOR GLYCOPROTEIN OSTEOPOINTIN SIALIC ACID CELL ADHESION SIGNAL BONE SIALOPROTEIN PD008016: E29-N273, M1-E271 | BLAST_PRODOM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|--|
| 5 | cont | | | | OSTEOPONTIN DM02611 P10451 1-313: E29-N273 P14287 1-302: E29-N273 P10923 1-293: S26-N273 P31098 1-276: D23-E231, P165-N273, M1-P96 Cell attachment sequence: R118-D120 Osteopontin Signature: K20-K30 signal_cleavage: M1-S19 | BLAST_DOMO |
| 6 | 90116002CD1 | 333 | S282 T22 T69 T133 T306 | | Signal Peptide: M1-G15 Signal Peptide: M1-S19 C1q (a subunit of the C1 enzyme complex that activates the serum complement system) domain: A203-L329 Collagen triple helix repeat (20 copies): R24-V82, A84-I142, G143-S202 C1q domain proteins BL01113: G149-P175, P220-I255, D285-Q304, T322-S331 Complement C1Q domain signature PR00007: F214-K240, F241-R260, D285-T306, D320-F330 PRECURSOR SIGNAL COLLAGEN REPEAT HYDROXYLATION GLYCOPROTEIN CHAIN PLASMA EXTRACELLULAR MATRIX PD002992: A203-L329 COLLAGEN ALPHA PRECURSOR CHAIN REPEAT SIGNAL CONNECTIVE TISSUE EXTRACELLULAR MATRIX PD000007: G89-G176, G101-E195, P40-D129, G26-E111 | MOTIFS MOTIFS SPSCAN HMMER HMMER HMMER_PFAM HMMER_PFAM BLIMPS_BLOCKS BLIMPS_PRINTS BLAST_PRODROM BLAST_PRODROM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|------------------------------------|-------------------------------|---|---|
| 6 | cont | | | | PRECURSOR SIGNAL COLLAGEN ALPHA 3(X) CHAIN EXTRACELLULAR MATRIX CONNECTIVE TISSUE PD028299: K85-G173, G32-G116, G107-G188, Q121-E185 SIMILAR TO CUTICULAR COLLAGEN PD067228: P61-K159, G107-E195, A9-G104+F75 CIQ DOMAIN DM00777 S23297 465-674: G116-L328 | BLAST_PRODROM BLAST_PRODROM BLAST_DOMO |
| | | | | | FIBRILLAR COLLAGEN CARBOXYL-TERMINAL DM00019 P12107 1270-1494: G26-G194 P20908 1300-1524: G26-G194 S18803 1304-1528: G26-G194 | BLAST_DOMO |
| | | | | | CIq domain signature: F223- Y253 | MOTIFS |
| 7 | 039283CD1 | 478 | S48 S68 S138 T3 T8 T45 T99 T131 | N17 N46 N66 N135 | OTOGELIN NASCENT POLYPEPTIDE ASSOCIATED COMPLEX ALPHA POLYPEPTIDE ALPHA NAC MUSCLE SPECIFIC FORM GP220 PD147940: P173-G477 FIBRILLAR COLLAGEN CARBOXYL-TERMINAL DM00019 S42886 221-377: P270-P432, P172-P326, P366-Y478 P08124 103-269: P276-A462, P172-S333, P321-G477 H-A-P-P (Extensin precursor (Cell wall hydroxyproline-rich glycoprotein) in tobacco & tomato) REPEAT DM08271 S25299 69-249: R290-W475, A170-P334 | BLAST_PRODROM BLAST_DOMO BLAST_DOMO |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|--|---|----------------------------------|
| 7 | cont | | | | ANONYMOUS; THROMBOSPONDIN (Thrombospondin related anonymous protein precursor); DM06934 Q01443 275-825: P178-P476, E155-P461 | BLAST_DOMO |
| 8 | 7505082CD1 | 185 | S17 S108 S129 S179 T39 Y48 | N154 | PROTOCOLADHERIN CELL ADHESION TRANSMEMBRANE CALCIUM-BINDING REPEAT GLYCOPROTEIN PROTEIN KIAA0588 PROTOCOLADHERIN 2 PD040979: Q61-A170 | BLAST_PRODOM |
| | | | | | PROTOCOLADHERIN CELL ADHESION TRANSMEMBRANE CALCIUM-BINDING REPEAT PROTOCOLADHERIN 2 PCDH2 COMPLETE CDS PD024559: V11-G60 signal_cleavage: M1-S22 | BLAST_PRODOM |
| 9 | 7505139CD1 | 170 | S140 T96 T141 | N26 N32 N41 N72 N77 N94 N104 N121 N127 | Signal Peptide: M1-C19, M1-V20, M1-A23, M1-K25, M1-T28, M1-D24, M1-S22 PUTATIVE MUCIN CORE PROTEIN 24 PRECURSOR MULTI-GLYCOSYLATED MGC24 MUC24 CD164 ANTIGEN GLYCOPROTEIN SIGNAL PD100758: M1-S140, G24-K170 Growth factor and cytokines receptors family signature 1: C70-W82 | SPSCAN |
| | | | | | Signal Peptide: M1-C19, M1-V20, M1-A23, M1-K25, M1-T28, M1-D24, M1-S22 | HMMER |
| | | | | | PUTATIVE MUCIN CORE PROTEIN 24 PRECURSOR MULTI-GLYCOSYLATED MGC24 MUC24 CD164 ANTIGEN GLYCOPROTEIN SIGNAL PD100758: M1-S140, G24-K170 | BLAST_PRODOM |
| | | | | | Growth factor and cytokines receptors family signature 1: C70-W82 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|---|----------------------------------|
| 10 | 7505234CD1 | 482 | S36 S267 S297 S371 S373 S386 S474 T215 T275 T293 T409 T415 T424 Y403 | N171 N420 | signal_cleavage: M1-A28 | SPSCAN |
| | | | | | Signal Peptide: K12-A28, M1-P32 | HMMER |
| | | | | | Cadherin domain: Y249-M340, I140-V235, L354-T442 | HMMER_PFAM |
| | | | | | Cadherins extracellular repeat proteins domain proteins BL00232: S330-P347, F79-L111, T227-G274 | BLIMPS_BLOCKS |
| | | | | | Cadherins extracellular repeated domain signature: T213-V263, I314-A368 | PROFILESAN |
| | | | | | CELL ADHESION TRANSMEMBRANE CALCIUM-BINDING REPEAT GLYCOPROTEIN KIAA0345-LIKE PROTO-CADHERIN PROTEIN PRECURSOR PD017893: A28-E139 | BLAST_PRODROM |
| | | | | | CADHERIN REPEAT DM00030 P39038 305-417: N276-D376 P33450 2417-2519: L196-D272 P10288 303-415: N276-D376 P33147 299-411: N276-D376 | BLAST_DOMO |
| | | | | | Cadherins extracellular repeated domain signature: L123-P133, I232-P242, V337-P347 | MOTIFS |
| 11 | 7500227CD1 | 117 | T39 T65 | | signal_cleavage: M1-G23 | SPSCAN |
| | | | | | Signal Peptide: M1-G18, M1-G20, M1-G23, M1-C30, M1-C24 | HMMER |
| | | | | | Leucine Rich Repeat: G77-M100, D53-P76 | HMMER_PFAM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|--------------------------------|--|----------------------------------|
| 11 | | | | | Leucine rich repeat N-terminal domain: G23-P51 | HMMER_PFAM |
| cont | | | | | | |
| 12 | 7503676CD1 | 485 | S100 S109 S236 S249 S421 T118 T210 T244 T286 T294 | N41 N79 N196 N398 N440 N446 | signal_cleavage: M1-A66 | SPSCAN |
| | | | | | Signal Peptide: M47-D63, M47-A64, M47-A66, M47-Q68, N41-A66, S40-A66, C39-A66, S42-A66 | HMMER |
| | | | | | F5/8 type C domain: T219-L370 | HMMER_PFAM |
| | | | | | Cytosolic domain: C474-N485 | TMHMMER |
| | | | | | Transmembrane domain: V451-V473 | |
| | | | | | Non-cytosolic domain: M1-D450 | |
| | | | | | Coagulation factors 5/8 type C domain proteins (FA58C) proteins BL01285: C216-S235 | BLIMPS_BLOCKS |
| | | | | | GLYCOPROTEIN PRECURSOR SIGNAL FACTOR REPEAT PROTEIN NEUROPHILIN CELL DOMAIN COAGULATION PD000875: L220-L370 | BLAST_PRODROM |
| | | | | | DISCOLDIN I N-TERMINAL DM00516 A42580 2085-2210: P252-C373 P12259 2095-2223: W250-C373 P21956 338-462: W250-C373 P00451 2221-2347: S249-Q374 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------------|-----------------------------|------------------------|---|----------------------------------|--|-------------------------------------|
| 13 | 7503606CD1 | 1489 | S79 S104 S174 S398 S478 S481 S591 S704 S763 S808 S813 S877 S947 S949 S977 S1012 S1045 S1058 S1104 S1162 S1285 S1305 S1329 S1422 S1426 S1460 S1468 S1475 T80 T177 T283 T368 T387 T565 T635 T698 T709 T733 T743 T747 T748 T859 T872 T886 T1087 T1123 T1134 T1153 T1181 T1201 T1205 T1234 T1304 Y394 Y625 | N215 N385 N989 N1230 | SAM domain (Sterile alpha motif): E528-S591, P597-A661 | HMMER PFAM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 13 | | | | | Ank repeat: N215-K247, E141-N173, T250-T279, S174-L204, P108-D140 | HMMER_PFAM |
| cont | | | | | NEGATIVE FACTOR F-PROTEIN PD00444: D954-G973, P107-M125, S139-E184 | BLIMPS_PRODOM |
| | | | | | PROTEIN REPEAT MICROTUBULE-ASSOCIATED MICROTUBULES PHOSPHORYLATION BASOON ALTERNATIVE SPLICING LARGE PROLINE-RICH PD005493: A1240-R1458, P1210-A1408 | BLAST_PRODOM |
| | | | | | PROTEIN TYROSINE PHOSPHATASE TD14 EC 3.1.3.48 HYDROLASE PD180360: P1246-G1456, P1252-D1453 | BLAST_PRODOM |
| | | | | | PROTEIN REPEAT SIGNAL PRECURSOR PRION GLYCOPROTEIN NUCLEAR GPI-ANCHOR BRAIN MAJOR PD001091: T1232-P1423, P1210-P1400 | BLAST_PRODOM |
| | | | | | DNA TOPOISOMERASE I ISOMERASE PD084780: P1254-P1414, G1231-A1410 | BLAST_PRODOM |
| | | | | | PROLINE-RICH PROTEIN DM03894 A39066 1-159: P1247-P1389 | BLAST_DOMO |
| | | | | | FORMIN; DM04565 Q05860 176-1467: G1231-P1372, R1238-P1400 | BLAST_DOMO |
| | | | | | H-A-P-P REPEAT DM0827 S25299 69-249: P1250-P1418, P1221-P1400 | BLAST_DOMO |
| | | | | | BAT2; DM05517 P48634 1-1860: D1052-R1428 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---|---|----------------------------------|
| 14 | 7500216CD1 | 703 | S90 S113 S452 S522 S679 T163 T208 T241 T400 T445 T657 Y162 Y646 | N44 N140 N198 N297 N308 N405 | Signal Peptide: M1-A25 | HMMER |
| | | | | | Signal Peptide: M1-A27 | HMMER |
| | | | | | Cytosolic domain: H549-I703 Transmembrane domain: M526-V548 Non-cytosolic domain: M1-D525 | TMHMMER |
| | | | | | Cadherins extracellular repeated domain signature: V112-P122 | MOTIFS |
| 15 | 7099880CD1 | 1032 | S81 S89 S102 S150 S165 S283 S374 S389 S425 S436 S444 S477 S492 S498 S513 S569 S683 S876 S890 S957 S981 S985 S992 T29 T167 T241 T277 T299 T387 T494 T543 T561 T595 T597 T658 T736 T794 T886 T927 T986 Y419 Y506 Y1018 | N48 N148 N306 N307 N347 N368 N450 N511 N630 N681 N734 N754 N775 N780 N901 | signal_cleavage: M1-A23 | SPSCAN |
| | | | | | Signal Peptide: M1-A21, M1-A23 | HMMER |
| | | | | | Cadherin domain: F577-M666, V257-T349, I474-L563, I147-S243, G369-E460, K685-Y773 | HMMER_PFAM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 15 | | | | | Cytosolic domain: R837-Y1032 Transmembrane domain: T814-V836 Non-cytosolic domain: M1-L813 | TMHMMER |
| | | | | | Cadherins extracellular repeated domain signature: V219-V271, T541-V591, I439-L488, V645-P693 | PROFILES CAN |
| | | | | | Cadherin signature PR00205: V267-G282, S656-P673, L274-R288 | BLIMPS_PRINTS |
| | | | | | CELL ADHESION TRANSMEMBRANE CALCIUM-BINDING REPEAT GLYCOPROTEIN PCDH7 NF PROTOCADHERIN BHPCDHA BHPCDHB PD014144: N775-S998 | BLAST_PRODOM |
| | | | | | CELL ADHESION TRANSMEMBRANE CALCIUM-BINDING REPEAT GLYCOPROTEIN KIAA0345-LIKE PROTOCADHERIN PROTEIN PRECURSOR PD017893: E25-S144 | BLAST_PRODOM |
| | | | | | INSECTICIDAL TOXIN RECEPTOR BTR1 PRECURSOR RECEPTOR GLYCOPROTEIN TRANSMEMBRANE SIGNAL REPEAT CELL ADHESION PD134331: I128-P570 | BLAST_PRODOM |
| | | | | | CADHERIN REPEAT DM00030 P33450 3259-3362:G172-D280 P33450 1079-1181:G499-D600 P33450 187-298: N174-I281 P33450 1308-1412: G499-D600 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|---|----------------------------------|
| 15 | cont | | | | Cadherins extracellular repeated domain signature: I130-P140, V240-P250, I346-P356, V457-P467, V560-P570, I663-P673 | MOTIFS |
| 16 | 871513CD1 | 687 | S116 S159 S255 S258 S278 S339 S351 S371 S409 S523 S557 S571 S606 S664 S673 T32 T70 T148 T228 T310 T388 T398 T479 T543 T613 T646 Y13 Y464 Y485 | N114 N337 N433 N671 | PH domain: P373-K476 | HMMER_PFAM |
| | | | | | C47E8.7 PROTEIN PD147334: L138-W686, D15-I178 | BLAST_PRODOM |

Table 3

| SEQ ID NO: | Incye Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|----------------------|---------------------|--|---|--|----------------------------------|
| 17 | 8057640CD1 | 1805 | S37 S128 S141 S202 S208 S282 S331 S368 S377 S439 S465 S496 S513 S672 S821 S893 S1002 S1025 S1140 S1151 S1167 S1212 S1239 S1339 S1360 S1364 S1526 S1530 S1752 T63 T90 T95 T386 T581 T612 T647 T882 T902 T1215 T1246 T1251 T1297 T1426 T1796 Y50 Y408 Y682 Y1657 | N327 N491 N579 N617 N695 N980 N1576 | Signal Peptide: M1-A27 | HMMER |
| | | | | | Fibronectin type III domain: P1511-R1595, P1624-R1711, L1127-S1208, P1220-R1310 | HMMER_PFAM |
| | | | | | Integrins, beta chain: S37-C455 | HMMER_PFAM |
| | | | | | Integrins beta chain cysteine-rich domain proteins BL00243: R60-M80, F118-R158, V159-W194, V206-H257, Q302-L345, G346-S387, R467-C492, L497-D539 | BLIMPS_BLOCKS |
| | | | | | Integrins beta chain cysteine-rich domain signature: N491-G548, G570-E660 | PROFILESAN |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 17 | cont | | | | INTEGRIN GLYCOPROTEIN CELL ADHESION TRANSMEMBRANE REPEAT PRECURSOR EXTRACELLULAR SUBUNIT SIGNAL PD001811: S37-D453 | BLAST_PRODUM |
| | | | | | INTEGRIN CELL ADHESION TRANSMEMBRANE GLYCOPROTEIN REPEAT BETA4INTEGRIN BETA4 SUBUNIT PRECURSOR PD011414: S710-L851 | BLAST_PRODUM |
| | | | | | PROTEIN EXCHANGER TRANSMEMBRANE GLYCOPROTEIN REPEAT PRECURSOR SIGNAL SODIUM/CALCIUM TRANSPORT NA+/CA2+EXCHANGE PD001766: N852-Q1094 | BLAST_PRODUM |
| | | | | | INTEGRIN CELL ADHESION TRANSMEMBRANE GLYCOPROTEIN REPEAT BETA4 INTEGRIN BETA4 SUBUNIT PRECURSOR PD011837: R1310-H1449, D1503- V1510 | BLAST_PRODUM |
| | | | | | INTEGRINS BETA CHAIN CYSTEINE-RICH DOMAIN DM00846 P16144 7-444: S7-K445 P32592 3-435: L12-L412, G1688-E1705 P09055 1-431: L14-L381 | BLAST_DOMO |
| | | | | | INTEGRINS BETA CHAIN CYSTEINE-RICH DOMAIN DM01094 P16144 618-787: Y618-D788 | BLAST_DOMO |
| | | | | | Cell attachment sequence: R1003-D1005 | MOTIFS |
| | | | | | EGF-like domain signature 1: C479-C490, C562- C573 | MOTIFS |
| | | | | | EGF-like domain signature 2: C479-C490, C562- C573 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---------------------------------|---|----------------------------------|
| 17 | | | | | Integrins beta chain cysteine-rich domain signature: C512-C525, C590-C603 | MOTIFS |
| 18 | 7505913CD1 | 372 | S45 S150 S237 S310 S364 T12 T92 T276 T296 | N107 N196 N308 | PH domain: P260-S348 | HMMER_PFAM |
| | | | | | Sec7 domain: T58-F243 | HMMER_PFAM |
| | | | | | PROTEIN ARF EXCHANGE FACTOR GUANINE NUCLEOTIDE RELEASING SEC7 CYTOHESIN TRANSPORT CHROMOSOME PD003358: K57-E241 | BLAST_PRODROM |
| | | | | | PROTEIN KINASE TRANSFERASE ATP BINDING SERINE/THREONINE PROTEIN MULTIGENE FAMILY SH3 DOMAIN RECEPTOR PD000761: R262-K336, K331-V347 | BLAST_PRODROM |
| | | | | | PROTEIN CYTOHESIN ARF NUCLEOTIDE BINDING SITE OPENER ARNO EXCHANGE FACTOR GUANINE NUCLEOTIDE PD150014: M17-Q60 | BLAST_PRODROM |
| 19 | 7510292CD1 | 1088 | S187 S188 S258 S268 S285 S415 S467 S547 S696 S810 S833 S865 S1019 S1029 S1040 S1083 T5 T35 T106 T198 T434 T483 T492 T925 T944 | N3 N490 N773 N960 N966 N1015 | signal_cleavage: M1-D24 | SPSCAN |
| | | | | | Signal Peptide: M1-C22 | HMMER |
| | | | | | Signal Peptide: M1-D24 | HMMER |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 19 | | | | | Signal Peptide: M1-S29 | HMMER |
| cont | | | | | Thrombospondin type 1 repeats W47-P102, W668-P724, Y726-G784, W786-T856, H861-E923, W928-Q985, R987-S1040 | HMMER_SMART |
| | | | | | PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS N PROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654; Q416-C484 | BLAST_PRODOM |
| | | | | | PROTEIN F25H8.3 F53B6.2 KIAA0605 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: W926-C1034, Y726-C855, W859-C984, W50-Q72, G941-C1039 | BLAST_PRODOM |
| | | | | | PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS N PROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161: R485-I599 | BLAST_PRODOM |
| | | | | | Leucine zipper pattern: L504-L525 | MOTIFS |
| 20 | 7504669CD1 | 436 | S220 S311 S315 S340 S347 S380 S397 T155 T385 Y105 | N75 N153 N237 N360 | signal_cleavage: M1-A22 | SPSCAN |
| | | | | | Signal Peptide: M1-A22 | HMMER |
| | | | | | Signal Peptide: M1-V25 | HMMER |
| | | | | | Signal Peptide: M1-S28 | HMMER |
| | | | | | Signal Peptide: M1-A26 | HMMER |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|---|---|----------------------------------|
| 20 | cont | | | | Signal Peptide: M1-E27 | HMMER |
| | | | | | PROCOLLAGEN LYSINE 2-OXOGLUTARATE 5-DIOXYGENASE PRECURSOR LYSYL HYDROXYLASE OXIDOREDUCTASE DIOXYGENASE SIGNAL IRON PD009947: P31-M270 | BLAST_PRODOM |
| | | | | | LYSYL HYDROXYLASE CHAIN DM07920 P24802 I-729: G24-D90. D139-Q278 | BLAST_DOMO |
| 21 | 7509266CD1 | 70 | T30 | N48 | signal_cleavage: M1-H18 | SPSCAN |
| | | | | | Signal Peptide: M1-G21 | HMMER |
| | | | | | Integrins, beta chain: T30-R58 | HMMER_PFAM |
| | | | | | PSI domain found in plexins, semaphorins, and integrins: G22-P68 | HMMER_SMART |
| | | | | | INTEGRINS BETA CHAIN CYSTEINE-RICH DOMAIN DM00846 P18564 I-442: M1-R58 | BLAST_DOMO |
| 22 | 7509288CD1 | 715 | S71 S103 S197 S224 S293 S300 S334 S391 S428 S439 S519 S553 S556 S584 S623 S662 T3 T78 T115 T116 T232 T505 T546 T582 T613 T695 T700 | N187 N314 N323 N390 N398 N468 N502 N698 | Integrins, beta chain: G43-C381 | HMMER_PFAM |
| | | | | | Integrin beta subunits (N-terminal portion) I2-C381 | HMMER_SMART |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 22 | | | | | Cytosolic domain: K658-C715 Transmembrane domain: I635-W657 Non-cytosolic domain: M1-N634 | TMHMMER |
| cont | | | | | Integrins beta chain cysteine-rich domain proteins BL00243: L48-K88, E89-C124, I140-H191, L193-L222, L234-V277, G278-A319, L478-C503, C518-C560 | BLIMPS_BLOCKS |
| | | | | | Integrins beta chain cysteine-rich domain signature: I459-S553 | PROFILES SCAN |
| | | | | | Integrins beta chain cysteine-rich domain signature: I459-E513 | PROFILES SCAN |
| | | | | | Integrins beta chain cysteine-rich domain signature: G414-G476 | PROFILES SCAN |
| | | | | | Type III EGF-like signature PR00011: G483-C501, V389-C417, G399-C417 | BLIMPS_PRINTS |
| | | | | | INTEGRIN GLYCOPROTEIN CELL ADHESION TRANSMEMBRANE REPEAT PRECURSOR EXTRACELLULAR SUBUNIT SIGNAL PD001811: I36-C381 | BLAST_PRODROM |
| | | | | | INTEGRIN GLYCOPROTEIN CELL ADHESION TRANSMEMBRANE REPEAT PRECURSOR EXTRACELLULAR SUBUNIT SIGNAL PD001794: T540-Y701 | BLAST_PRODROM |
| | | | | | INTEGRIN GLYCOPROTEIN CELL ADHESION TRANSMEMBRANE REPEAT SUBUNIT PRECURSOR EXTRACELLULAR MATRIX PD149771: D382-Y463 | BLAST_PRODROM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|---|----------------------------------|
| 22 | cont | | | | INTEGRIN CELL ADHESION TRANSMEMBRANE GLYCOPROTEIN REPEAT PRECURSOR SIGNAL BETA G PD154651: N390-Y463 INTEGRINS BETA CHAIN CYSTEINE-RICH DOMAIN DM00846 P18564 I-442: L38-A370 DM00846 P18084 7-451: I36-A370 DM00846 P05106 9-449: I36-A370 DM00846 JC2005 I-308: M66-A370 Cell attachment sequence: R441-D443, R521-D523 | BLAST_PRODOM |
| | | | | | Cytochrome c family heme-binding site signature: C557-L562 | MOTIFS |
| | | | | | EGF-like domain signature 1: C406-C417, C490-C501 | MOTIFS |
| | | | | | EGF-like domain signature 2: C490-C501 | MOTIFS |
| 23 | 7510212CD1 | 596 | S7 S71 S213 S226 S239 S259 S285 S379 S438 S442 S468 S502 S525 S567 T161 T193 T250 T508 | N98 N565 N569 | signal_cleavage: M1-A29 | SPSCAN |
| | | | | | Signal Peptide: M1-A25 | HMMER |
| | | | | | Signal Peptide: M1-A29 | HMMER |
| | | | | | Anaphylotoxin-like domain: C36-C69, C112-C144, H77-C110 | HMMER_PFAM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 23 | cont | | | | EGF-like domain: C360-C397, C515-C553, C403-C439, C312-C354, C180-C214, C475-C509, E440-C469, C266-C306 | HMMER_PFAM |
| | | | | | Calcium-binding EGF-like domain C180-E215, D216-K261, D262-I307, D308-V355, D356-V398, D399-E440, S442-E470, D471-E510, D511-Q554, D555-T596 | HMMER_SMART |
| | | | | | Epidermal growth factor-like domain R179-E215, E219-K261, E265-I307, E311-V355, E359-V398, E402-E440, S442-E470, E474-E510, E514-Q554, E558-X597 | HMMER_SMART |
| | | | | | Anaphylatoxin homologous domain C36-C69, H77-C110, C112-C144 | HMMER_SMART |
| | | | | | Anaphylatoxin domain proteins BL01177: G18-A29, S200-E215, S478-R496, V504-G521, Y526-N552 | BLIMPS_BLOCKS |
| | | | | | Calcium-binding EGF-like domain proteins pattern proteins BL01187: C469-C480, C415-F430 | BLIMPS_BLOCKS |
| | | | | | Type II EGF-like signature PR00010: D471-Q482, G183-C190, G420-F430 | BLIMPS_PRINTS |
| | | | | | GLYCOPROTEIN EGF-LIKE DOMAIN PRECURSOR FIBULIN I ISOFORM SIGNAL EXTRACELLULAR MATRIX PLASMA PD006208: V27-N177 | BLAST_PRODROM |
| | | | | | FIBULIN2 PRECURSOR SIGNAL GLYCOPROTEIN EXTRACELLULAR MATRIX PLASMA EGF-LIKE DOMAIN CALCIUM-BINDING ALTERNATIVE SPLICING REPEAT PD166194: C186-C227 | BLAST_PRODROM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 23 | cont | | | | EGF DM00003 P98095 720-782: V217-N281 DM00003 P98095 860-919: D358-T418 DM00003 P98163 1373-1460: C426-I512, R353-E440 | BLAST_DOMO |
| | | | | | EGF-LIKE DOMAIN DM00864 I55476 159-241: C480-C559, S442-C515, C325-R405, R181-E267, C373-E446 | BLAST_DOMO |
| | | | | | Anaphylatoxin domain signature: C36-C69, H77-C110, C112-C144 | MOTIFS |
| | | | | | Aspartic acid and asparagine hydroxylation site: C373-C384, C415-C426, C445-C456, C484-C495, C528-C539 | MOTIFS |
| | | | | | EGF-like domain signature 2: C199-C214, C382-C397, C424-C439, C454-C469 | MOTIFS |
| | | | | | Calcium-binding EGF-like domain pattern signature: D216-C242, D262-C288, D308-C334, D356-C382, D399-C424, D471-C493, D511-C537, D555-C581 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|---|---|----------------------------------|
| 24 | 7510504CD1 | I762 | S22 S28 S56 S62 S77 S120 S252 S329 S402 S414 S475 S558 S574 S651 S763 S803 S840 S856 S973 S976 S1006 S1019 S1054 S1111 S1123 S1128 S1144 S1149 S1157 S1171 S1177 S1224 S1344 S1352 S1463 S1689 T8 T25 T169 T184 T199 T235 T320 T413 T423 T650 T873 T994 T1052 T1053 T1165 T1275 T1283 T1295 T1378 T1567 T1571 T1699 T1723 T1733 Y226 Y945 | N251 N702 N828 N1004 N1051 N1084 N1251 N1303 N1323 N1342 N1427 N1458 N1518 | signal_cleavage: M1-S28 | SPSCAN |
| | | | | | Signal Peptide: M1-S23 | HMMER |
| | | | | | Signal Peptide: M1-A26 | HMMER |
| | | | | | Signal Peptide: M1-S28 | HMMER |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 24 cont | | | | | Immunoglobulin domain: G1301-A1355, K892-A949, G1411-A1471, S1197-A1252, S1547-C1607, V1688-C1725 | HMMER_PFAM |
| | | | | | Thrombospondin type 1 domain: D37-C81, F526-C583, S1547-C1607, A735-C788, S790-C849, W670-P717, W440-C492, W380-C437, V1668-C1725, W611-C666 | HMMER_PFAM |
| | | | | | Thrombospondin type 1 repeats: W36-P82, W301-P360, R379-P438, K439-Y493, S525-S584, D610-P667, R669-P729, W732-P789, E791-A850, W1490-P1546, W1549-V1608, E1609-S1667, H1669-E1726 | HMMER_SMART |
| | | | | | Immunoglobulin C-2 type (IgC2) L890-E954, A1192-G1257, V1299-G1360, D1409-G1476 | HMMER_SMART |
| | | | | | Immunoglobulin G884-R970, Q1187-A1268, G1293-L1371, G1403-Q1487 | HMMER_SMART |
| | | | | | Ig Superfamily from SCOP Q872-A952, V1181-V1273, V1288-V1376, V1397-L1478 | HMMER_INCYTE |
| | | | | | I type Ig domains from SCOP A1178-K1274, E1284-P1377 | HMMER_INCYTE |
| | | | | | RECEPTOR INTERLEUKIN-1 PD02870: L1235-L1267, P1714-L1748, L1338-D1372, V1397-K1421 | BLIMPS_PRODROM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|---|----------------------------------|
| 24 | cont | | | | PROTEIN F25H8.3 F53B6.2 KIAA0605 PROCOLLAGEN C37C3.6 SERINE PROTEASE INHIBITOR ALTERNATIVE PD007018: W670-E791, W1490-C1607, R300-P438, W732-C844, W611-P730 PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS N PROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654: P115-C185 | BLAST_PRODOM |
| 25 | 7510587CD1 | 438 | S25 S203 S224 S324 S345 S358 S402 S425 T183 T398 Y34 | N38 N175 N314 N360 | Papain family cysteine protease: L217-G417 | HMMER_PFAM |
| | | | | | Somatomedin B domain: N59-K106 | HMMER_SMART |
| | | | | | Eukaryotic thiol (cysteine) proteases cysteine proteins BL00139: Q237-F246, R281-I289, T410-L419 | BLIMPS_BLOCKS |
| | | | | | Somatomedin B signature PR00022: A79-E90, S92-R103 | BLIMPS_PRINTS |
| | | | | | Papain cysteine protease (C1) family signature PR00705: Q237-A252 | BLIMPS_PRINTS |
| | | | | | F26E4.3 PROTEIN TUBULOINTERSTITIAL NEPHRITIS ANTIGEN PD037748: G54-M207 | BLAST_PRODOM |
| | | | | | TUBULOINTERSTITIAL NEPHRITIS ANTIGEN PD113487: M1-Q53 | BLAST_PRODOM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|-------------------------------|--|----------------------------------|
| 25 | cont | | | | EUKARYOTIC THIOL (CYSTEINE) PROTEASES CYSTEINE DM00081 A57480 83-465: K85-R429 DM00081 S52212 34-335: L155-G417 DM00081 P07688 20-329: E159-T396 DM00081 P43157 24-339: S150-T398 | BLAST_DOMO |
| 26 | 7510684CD1 | 401 | S25 S128 S149 S249 S270 S283 S327 S366 S382 S397 T108 T323 T345 T396 Y34 Y372 | N38 N100 N239 N285 N380 | Papain family cysteine protease: L142-W392 | HMMER_PFAM |
| | | | | | Eukaryotic thiol (cysteine) proteases cysteine proteins BL00139: Q162-F171, R206-I214, T335-G344, F356-Y372 | BLIMPS_BLOCKS |
| | | | | | Papain cysteine protease (C1) family signature PR00705: Q162-A177, H336-L346, F356-S362 | BLIMPS_PRINTS |
| | | | | | F26E4.3 PROTEIN TUBULOINTERSTITIAL NEPHRITIS ANTIGEN PD037748: Q43-M132 | BLAST_PRODOM |
| | | | | | PROTEASE PRECURSOR SIGNAL CYSTEINE PROTEINASE HYDROLASE THIOL ZYMOGEN CATHEPSIN GLYCOPROTEIN PD000158: D161-R377 | BLAST_PRODOM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---------------------------------|--|----------------------------------|
| 26 | cont | | | | EUKARYOTIC THIOL (CYSTEINE) PROTEASES CYSTEINE DM00081 A57480 83-465: Q43-G393 DM00081 S52212 34-335: L80-A390 DM00081 P25792 23-338: E84-A390 DM00081 S31909 1-313: E84-A390 ATP/GTP-binding site motif A (P-loop): A359-S366 | BLAST_DOMO |
| | | | | | Eukaryotic thiol (cysteine) proteases asparagine active site: F356-I375 | MOTIFS |
| 27 | 7510697CD1 | 1074 | S187 S188 S258 S268 S285 S415 S467 S547 S696 S796 S819 S851 S1005 S1015 S1026 S1069 T5 T35 T106 T198 T434 T483 T492 T911 T930 | N3 N490 N773 N946 N952 N1001 | signal_cleavage: M1-D24 | SPSCAN |
| | | | | | Signal Peptide: M1-C22 | HMMER |
| | | | | | Signal Peptide: M1-D24 | HMMER |
| | | | | | Signal Peptide: M1-S29 | HMMER |
| | | | | | Thrombospondin type 1 repeats W47-P102, W668-P724, Y726-G784, W786-T842, H847-E909, W914-Q971, R973-S1026 | HMMER_SMART |
| | | | | | PRECURSOR GLYCOPROTEIN PD01719: W786-N813, R1018-C1025 | BLIMPS_PRODOM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 27 cont | | | | | PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS N PROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD011654; Q416-C484 | BLAST_PRODUM |
| | | | | | PROTEIN PROCOLLAGEN THROMBOSPONDIN MOTIFS N PROTEINASE A DISINTEGRIN METALLOPROTEASE WITH ADAMTS1 PD014161; R485-I599 | BLAST_PRODUM |
| | | | | | Leucine zipper pattern: L504-L525 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---|---|----------------------------------|
| 28 | 7761337CD1 | 1564 | S35 S86 S124 S135 S162 S186 S373 S506 S606 S616 S705 S722 S760 S767 S807 S875 S903 S922 S931 S1075 S1092 S1102 S1252 S1304 S1454 S1486 S1502 T40 T101 T152 T212 T262 T306 T337 T399 T430 T492 T523 T585 T726 T799 T800 T801 T847 T852 T867 T891 T947 T983 T987 T1160 T1174 T1218 T1227 T1243 T1248 T1257 T1269 T1290 T1315 T1472 Y768 Y1400 | N38 N166 N184 N327 N788 N1018 N1034 N1172 N1525 | signal_cleavage: M1-A19 | SPSCAN |
| | | | | | Signal Peptide: M1-A19 | HMMER |
| | | | | | Signal Peptide: M1-E21 | HMMER |
| | | | | | Signal Peptide: M1-G22 | HMMER |
| | | | | | Signal Peptide: M1-V24 | HMMER |
| | | | | | Signal Peptide: M4-G22 | HMMER |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 28 cont | | | | | EGF-like domain: C408-C434, C501-C527, C377-C403, C563-C589, C470-C496, C284-C310, C532-C558, C594-C620, C315-C341, C185-C216, C252-C279, C346-C372, C221-C247, C439-C465 Fibrinogen beta and gamma chains, C-term: F1343-S1552 | HMMER_PFAM |
| | | | | | Fibronectin type III domain L1250-S1327, L893-S974, M1073-S1151, L985-S1062, L803-S882, V623-S701, L1162-S1239, L712-L795 | HMMER_PFAM |
| | | | | | Fibronectin type III domain V623-P703, L712-G794, L803-S882, L893-S974, L985-S1062, M1073-S1151, L1162-S1239, L1250-S1327 | HMMER_SMART |
| | | | | | Fibrinogen-related domains (FreDs) P1342-S1552 | HMMER_SMART |
| | | | | | Epidermal growth factor-like domain F160-S186, E189-S217, A220-S248, I251-N280, L283-S311, I314-G342, T345-S373, R376-G404, K407-S435, R438-S466, S469-R497, Q500-A528, S531-K559, R562-G590, S593-S621 | HMMER_SMART |
| | | | | | Cytosolic domain: M1-Q6 Transmembrane domain: L7-L25 Non-cytosolic domain: K26-A1564 | TMHMMER |
| | | | | | Fibrinogen beta and gamma chains C-terminal domain proteins BL00514: V1375-G1411, E1416-T1428, Y1462-S1476, N1498-S1527, S1527-P1551 | BLIMPS_BLOCKS |
| | | | | | Fibrinogen beta and gamma chains C-terminal domain signature: D1489-E1539 | PROFILES SCAN |

Table 3

| SEQ ID NO: | Incye Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 28 | | | | | Type III EGF-like signature PR00011: G571-C589, G540-C558 | BLIMPS_PRINTS |
| cont | | | | | Glycoprotein Matrix E1 PD01521 A1446-A1458 | BLIMPS_PRODROM |
| | | | | | PRECURSOR GLYCOPROTEIN SIGNAL | BLAST_PRODROM |
| | | | | | FIBRINOGEN BLOOD COAGULATION CHAIN | |
| | | | | | PLASMA PROTEIN PLATELET | |
| | | | | | PD001241: K1345-S1541, S1327-P1551, T306-I314 | |
| | | | | | GLYCOPROTEIN PRECURSOR TENASCIN | BLAST_PRODROM |
| | | | | | SIGNAL MATRIX CYTOTACTIN ANTIGEN | |
| | | | | | CELL TENASCIN X EGF-LIKE | |
| | | | | | PD004440: M4-R137 | |
| | | | | | PROTEIN TRANSCRIPTIONAL REPEAT | BLAST_PRODROM |
| | | | | | TRANSCRIPTION REGULATION DNA-BINDING | |
| | | | | | NUCLEAR SHUTTLE CRAFT PUTATIVE | |
| | | | | | PD014613: C161-P625 | |
| | | | | | REELIN GLYCOPROTEIN | BLAST_PRODROM |
| | | | | | PD150229: C408-K453, C346-G416, C470-C496, | |
| | | | | | C501-C527, C563-S596, C284-A349, C190-S217, | |
| | | | | | F160-C185, C221-S248, P188-C216, D302-T345, | |
| | | | | | P282-C310, C221-C247 | |
| | | | | | FIBRINOGEN BETA/GAMMA | BLAST_DOMO |
| | | | | | DM00531 P24821 1946-2187: P1310-N1553 | |
| | | | | | DM00531 S19694 1492-1734: P1310-N1553 | |
| | | | | | DM00531 P10039 1554-1796: P1310-N1553 | |
| | | | | | DM00531 JH0675 1096-1338: S1311-N1553 | |
| | | | | | Cell attachment sequence: R877-D879 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|---------------------------------|--|----------------------------------|
| 28 | cont | | | | EGF-like domain signature 1: C174-C185, C205-C216, C236-C247, C268-C279, C299-C310, C330-C341, C361-C372, C392-C403, C423-C434, C454-C465, C485-C496, C516-C527, C547-C558, C578-C589, C609-C620 | MOTIFS |
| | | | | | EGF-like domain signature 2: C174-C185, C205-C216, C236-C247, C268-C279, C299-C310, C330-C341, C361-C372, C392-C403, C423-C434, C454-C465, C485-C496, C516-C527, C547-C558, C578-C589, C609-C620 | MOTIFS |
| 29 | 7503666CD1 | 834 | S90 S113 S441 S653 S691 S810 T163 T208 T241 T389 T434 T466 T498 T531 T788 Y162 Y777 | N44 N140 N198 N297 N394 N515 | Signal Peptide: M1-A25, M1-A27 | HMMER |
| | | | | | Cadherin repeats.: Y29-P122, L149-P235 | HMMER_PFAM |
| | | | | | Cytosolic domain: H680-I834 | TMHMMER |
| | | | | | Transmembrane domain: M657-V679 | |
| | | | | | Non-cytosolic domain: M1-D656 | |
| | | | | | Cadherins extracellular repeat proteins domain proteins BL00232: Q107-R154 | BLIMPS_BLOCKS |
| | | | | | MYROSINASE-BINDING PROTEIN PD033908: P501-P646, E735-G757 | BLAST_PRODROM |
| | | | | | LECTIN ANTIGEN IGE-BINDING REPEAT ANNEXIN GALECTIN-3 GALACTOSE-SPECIFIC MAC-2 CBP GALAPTIN PD001091: E442-S653 | BLAST_PRODROM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|---|--|----------------------------------|
| 29 | | | | | PROLINE-RICH PROTEIN DM03894 A39066 1-159; P493-P627, Q439-P573 | BLAST_DOMO |
| cont | | | | | EPSTEIN;ANTIGEN;BARR; MEMBRANE; DM06222 B43042 1-710: T426-S641 | BLAST_DOMO |
| | | | | | PHASE; ACUTE; SHED; ANTIGEN; DM05630 P23253 951-1161: P462-H680, T428-T610 | BLAST_DOMO |
| | | | | | Cadherins extracellular repeated domain signature: V112-P122 | MOTIFS |
| 30 | 7503668CD1 | 814 | S90 S421 S633 S671 S790 T132 T177 T210 T369 T414 T446 T478 T511 T768 Y131 Y757 | N44 N109 N167 N266 N277 N374 N495 | Signal Peptide: M1-A25, M1-A27 | HMMER |
| | | | | | Cytosolic domain: H660-I814 Transmembrane domain: M637-V659 Non-cytosolic domain: M1-D636 | TMHMMER |
| | | | | | MYROSINASE-BINDING PROTEIN PD033908: P481-P626, E715-G737 | BLAST_PRODROM |
| | | | | | LECTIN ANTIGEN IGE-BINDING REPEAT ANNEXIN GALECTIN-3 GALACTOSE-SPECIFIC MAC-2 CBP GALAPTIN PD001091: E422-S633 | BLAST_PRODROM |
| | | | | | PROLINE-RICH PROTEIN DM03894 A39066 1-159; Q419-P553, R452-P538, P473-P607 | BLAST_DOMO |
| | | | | | EPSTEIN;ANTIGEN;BARR; MEMBRANE; DM06222 B43042 1-710: T406-S621 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|--|---|---|----------------------------------|
| 30 cont | | | | | PHASE; ACUTE; SHED; ANTIGEN; DM05630 P23253 951-1161: P442-H660, T408-T590 | BLAST_DOMO |
| 31 | 7503672CD1 | 807 | S90 S113 S452 S626 S664 S783 T163 T208 T241 T400 T445 T477 T509 T542 T761 Y162 Y750 | N44 N140 N198 N297 N308 N405 N526 | Signal Peptide: M1-A25, M1-A27 | HMIMER |
| | | | | | Cadherin repeats.: Y29-P122, L149-P235, I274-P352 | HMIMER_PFAM |
| | | | | | Cytosolic domain: H653-I807 Transmembrane domain: M630-V652 Non-cytosolic domain: M1-D629 | TMHMMER |
| | | | | | Cadherins extracellular repeat proteins domain proteins BL00232: Q107-R154 | BLIMPS_BLOCKS |
| | | | | | PROLINE-RICH PROTEIN | BLAST_DOMO |
| | | | | | DM03894 A39066 [-159: Q450-P584, R483-E621 | BLAST_DOMO |
| | | | | | PHASE; ACUTE; SHED; ANTIGEN; DM05630 | BLAST_DOMO |
| | | | | | P23253 951-1161: T439-H653 S23006 666-879: A467-H653 | |
| | | | | | Cadherins extracellular repeated domain signature: V112-P122 | MOTIFS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|-------------------------------|--|----------------------------------|
| 32 | 6039650CD1 | 1232 | S10 S28 S34 S42 S133 S138 S156 S166 S190 S234 S239 S244 S338 S403 S448 S485 S530 S596 S601 S612 S697 S725 S745 S763 S789 S833 S838 S915 S1071 S1163 S1169 T8 T53 T199 T215 T281 T289 T323 T381 T387 T493 T572 T701 T735 T822 T952 T963 T976 T1028 T1152 T1219 T1226 | N119 N772 N788 N837 N1215 | Sterile alpha motif: F875-L944, G990-L1057, V1078-M1150 | HMMER SMART |
| | | | | | SAM domain (Sterile alpha motif): A876-M942, D991-R1055 | HMMER PFAM |
| | | | | | PROTEIN LAR-INTERACTING F59F5.6 LIPRIN ALPHA2 1A 1B PD025515: S28-K115; L191-N211; L378-N440; D475-L865; G621-T956 | BLAST_PRODOM |
| | | | | | PROTEIN LAR-INTERACTING LIPRIN ALPHA2 1A 1B F59F5.6 PD023244: Q29-I350 | BLAST_PRODOM |
| | | | | | PROTEIN LAR-INTERACTING COILED-COIL LIKE LIPRIN BETA1 T21H8.1 LIPRIN ALPHA2 1A 1B F59F5.6 PD014510: Y989-D1160 | BLAST_PRODOM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 32 | cont | | | | LAR-INTERACTING PROTEIN LIPRIN ALPHA2 1A 1B PD043722: D1161-G1214 | BLAST_PRODUM |
| | | | | | TRICHOHYALIN DM03839 P37709 632-1103: E37-R461; E254-E658; E853-L873; L1054-R1088; E1141-R1175 | BLAST_DOMO |
| | | | | | MYOSIN-LIKE PROTEIN MLP-1 DM07884 Q02455 35-1728: E46-N477 | BLAST_DOMO |
| | | | | | TRICHOHYALIN DM03839 P22793 921-1475: S28-Q499; E102-R663; R1055-D1077 | BLAST_DOMO |
| | | | | | TRICHOHYALIN DM03839 Q07283 91-443: E226-Q489 | BLAST_DOMO |
| 33 | 7509919CD1 | 132 | S32 S111 S119 | | signal_cleavage: M1-S22 Signal Peptide: M1-S22; M1-E24; M1-S32 | SPSCAN HMMER |
| | | | | | CELL SURFACE ADHESION GLYCOPROTEINS LFA1/CR3/P150,95 BETASUBUNIT PRECURSOR INTEGRIN BETA2 CD18 ANTIGEN COMPLEMENT RECEPTOR C3 TRANSMEMBRANE GLYCOPROTEIN REPEAT EXTRACELLULAR MATRIX PD162030: M1-S31 | BLAST_PRODUM |
| | | | | | INTEGRINS BETA CHAIN CYSTEINE-RICH DOMAIN DM00846 P32592 3-435: R5-L49; Q47-V66 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incye Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|----------------------|---------------------|---|---|--|----------------------------------|
| 34 | 7510758CD1 | 886 | S44 S78 S80 S121 S173 S217 S245 S328 S381 S398 S457 S476 S544 S631 S649 S679 S828 S862 T49 T92 T139 T192 T230 T265 T387 T452 T511 T516 T663 T808 T884 Y212 | N42 N90 N235 N247 N257 N307 N331 N432 N655 N747 N826 | signal_cleavage: M1-G18 | SPSCAN |
| | | | | | Signal Peptide: M1-G18; M1-G20; M1-A23 | HMMER |
| | | | | | I type Ig domains from SCOP: A238-T333 | HMMER_INCY |
| | | | | | Ig superfamily from SCOP: D41-P117, Q138-A241, K243-N331, P338-V424, V444-P538, V541-A624, | HMMER_INCY |
| | | | | | Immunoglobulin: E45-Q127, D142-T236, N247-R327, K342-S434, R448-Q536, S545-K633 | HMMER_SMART |
| | | | | | Immunoglobulin C-2 Type: R51-G115, Y148-V221, N253-G315, Q348-A422, R454-G521, A551-G622 | HMMER_SMART |
| | | | | | Domain in meprin, A5, receptor protein tyrosine phosphatase mu (and others): L748-Y881 | HMMER_SMART |
| | | | | | MAM domain.: C753-Y881 | HMMER_PFAM |
| | | | | | Immunoglobulin domain: G53-A110, E150-V216, G255-A310, G350-A417, G456-T516, G553-V617 | HMMER_PFAM |
| | | | | | MAM domain proteins BL00740: C761-W773 | BLIMPS_BLOCKS |
| | | | | | MAM domain signature PR00020: K759-N777, Y832-K843 | BLIMPS_PRINTS |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---|--|---|----------------------------------|
| 34 | cont | | | | <p>PRECURSOR GLYCOPROTEIN SIGNAL</p> <p>TRANSMEMBRANE HYDROLASE PROTEIN</p> <p>REPEAT RECEPTOR PHOSPHATASE</p> <p>NEUROFILIN PD001482: D750-G871</p> <p>HEMICENTIN PRECURSOR SIGNAL</p> <p>GLYCOPROTEIN EGFLIKE DOMAIN HIM4</p> <p>PROTEIN ALTERNATIVE SPLICING PD066634: P240-C415</p> | BLAST_PRODUM |
| 35 | 7510063CD1 | 859 | <p>S27 S32 S104 S344</p> <p>S349 S472 S480</p> <p>S627 S665 S781</p> <p>T99 T125 T186</p> <p>T273 T311 T317</p> <p>T340 T357 T381</p> <p>T794 T817 Y187</p> <p>Y285</p> | <p>N222 N315 N347</p> <p>N433 N459 N488</p> <p>N811 N855</p> | <p>signal_cleavage: M1-S19</p> | SPSCAN |
| | | | | | Signal Peptide: M1-S19 | HMMER |
| | | | | | I type Ig domains from SCOP: S19-K121, V210-A306, K307-Q419, G420-S514 | HMMER_INCY |
| | | | | | Ig superfamily from SCOP: V22-L118, F120-I214, A216-K309, I310-A415, K417-E498 | HMMER_INCY |
| | | | | | Fibronectin type 3 domain: P510-G595, P612-S694 | HMMER_SMART |
| | | | | | Immunoglobulin: P26-F115, P124-N209, I220-F305, E314-Q413, P421-D508 | HMMER_SMART |
| | | | | | Immunoglobulin C-2 Type: S32-G103, R130-R196, N226-G294, E320-G402, W427-G496 | HMMER_SMART |
| | | | | | Fibronectin type III domain: P510-S598, P612-A697 | HMMER_PFAM |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 35 | | | | | Immunoglobulin domain: G34-V98, G132-G191, G228-A289, E322-A397, G429-A491 | HMMER_PFAM |
| cont | | | | | Non-cytosolic domain: M1-A724 Transmembrane domain: I725-L747 Cytosolic domain: N748-A859 | TMHMMER |
| | | | | | Fibronectin type III repeat signature PR00014: P442-D451, G639-Y649, L663-Y681, Y681-K695 | BLIMPS_PRINTS |
| | | | | | GLYCOPROTEIN ANTIGEN PRECURSOR PD02327: L302-V313, I434-L455, S704-S718 | BLIMPS_PRODOM |
| | | | | | NEURAL CELL ADHESION PRECURSOR NCAM GLYCOPROTEIN TRANSMEMBRANE REPEAT IMMUNOGLOBULIN FOLD PD006197: D741-G805 | BLAST_PRODOM |
| | | | | | NEURAL CELL ADHESION PRECURSOR NCAM GLYCOPROTEIN TRANSMEMBRANE REPEAT IMMUNOGLOBULIN FOLD PD008310: P810-A859 | BLAST_PRODOM |
| | | | | | NEURAL CELL ADHESION ISOFORM PRECURSOR NCAM GLYCOPROTEIN TRANSMEMBRANE REPEAT IMMUNOGLOBULIN FOLD PD009498: A697-V740 | BLAST_PRODOM |
| | | | | | NEURAL CELL ADHESION MOLECULE NCAM CELL ADHESION TRANSMEMBRANE PD000496: S566-P612 | BLAST_PRODOM |
| | | | | | NEURAL CELL ADHESION MOLECULE DM03574 P31836 735-852: V738-A859 | BLAST_DOMO |
| | | | | | IMMUNOGLOBULIN DM00001 P31836 310-404: T311-Q406 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incye Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 35 | | | | | IMMUNOGLOBULIN DM00001 P31836 22-112; V22-K113 | BLAST_DOMO |
| | | | | | IMMUNOGLOBULIN DM00001 P13594 22-112; V22-K113 | BLAST_DOMO |
| 36 | 7510135CD1 | 195 | S37 S42 S61 S93 S127 S130 T94 | N74 | signal_cleavage: M1-A30 | SPSCAN |
| | | | | | Signal Peptide: M1-A30 | HMMER |
| | | | | | Integrin alpha (beta-propellor repeats): E45-A104 | HMMER_SMART |
| | | | | | FG-GAP repeat: G46-R108 | HMMER_PFAM |
| | | | | | Integrins alpha chain proteins BL00242: E82-S93 | BLIMPS_BLOCKS |
| | | | | | INTEGRIN PRECURSOR CELL GLYCOPROTEIN SIGNAL ADHESION TRANSMEMBRANE EXTRACELLULAR MATRIX CYTOSKELETON PD001587: A30-R145 | BLAST_PRODROM |
| | | | | | INTEGRIN SUBUNIT VITRONECTIN RECEPTOR ALPHA PRECURSOR ALPHA-V CELL ADHESION GLYCOPROTEIN PD150841: T146-S174 | BLAST_PRODROM |
| | | | | | INTEGRINS ALPHA CHAIN DM00458 P26008 28-213: A39-K195 | BLAST_DOMO |
| | | | | | INTEGRINS ALPHA CHAIN DM00458 S60571 38-224: Y41-K195 | BLAST_DOMO |
| | | | | | INTEGRINS ALPHA CHAIN DM00458 P26009 32-222: Y41-K195 | BLAST_DOMO |
| | | | | | INTEGRINS ALPHA CHAIN DM00458 P53708 9-202: A39-K195 | BLAST_DOMO |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 37 | | | | | Signal Peptide: M1-A19; M1-E21; M1-G22; M1-V24; M4-G22 | HMMER |
| cont | | | | | Epidermal growth factor-like domain.: F160-S186; E189-S217; A220-S248; I251-N280; L283-S311; I314-G342; T345-S373; R376-G404; K407-S435; R438-S466; S469-R497; Q500-A528; S531-K559; R562-G590; S593-S621 | HMMER_SMART |
| | | | | | Fibrinogen-related domains (FReDs): P1888-S2098 | HMMER_SMART |
| | | | | | Fibronectin type 3 domain: V623-P703; L712-G794; L803-S882; L893-S974; L985-S1062; R1073-R1152; T1165-T1245; V1256-S1334; L1347-R1425; E1438-K1518; E1529-S1608; M1619-S1697; L1708-S1785; L1796-S1873 | HMMER_SMART |
| | | | | | EGF-like domain: C185-C216; C221-C247; C252-C279; C284-C310; C315-C341; C346-C372; C377-C403; C408-C434; C439-C465; C470-C496; C501-C527; C532-C558; C563-C589; C594-C620 | HMMER_PFAM |
| | | | | | Fibrinogen beta and gamma chains, C-terminal globular domain: F1889-S2098 | HMMER_PFAM |
| | | | | | Fibronectin type III domain: V623-S701; L712-L795; L803-S882; L893-S974; L985-S1062; A1074-S1157; T1165-S1243; V1256-T1335; L1347-S1430; E1438-S1514; E1529-S1608; M1619-S1697; L1708-S1785; L1796-S1873 | HMMER_PFAM |
| | | | | | Cytosolic domain: M1-Q6 Transmembrane domain: L7-L25 Non-cytosolic domain: K26-A2110 | TMHMMER |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|---|----------------------------------|
| 37 | | | | | Signal Peptide: M1-A19; M1-E21; M1-G22; M1-V24; M4-G22 | HMMER |
| cont | | | | | Epidermal growth factor-like domain.: F160-S186; E189-S217; A220-S248; I251-N280; L283-S311; I314-G342; T345-S373; R376-G404; K407-S435; R438-S466; S469-R497; Q500-A528; S531-K559; R562-G590; S593-S621 | HMMER_SMART |
| | | | | | Fibrinogen-related domains (FReDs): P1888-S2098 | HMMER_SMART |
| | | | | | Fibronectin type 3 domain: V623-P703; L712-G794; L803-S882; L893-S974; L985-S1062; R1073-R1152; T1165-T1245; V1256-S1334; L1347-R1425; E1438-K1518; E1529-S1608; M1619-S1697; L1708-S1785; L1796-S1873 | HMMER_SMART |
| | | | | | EGF-like domain: C185-C216; C221-C247; C252-C279; C284-C310; C315-C341; C346-C372; C377-C403; C408-C434; C439-C465; C470-C496; C501-C527; C532-C558; C563-C589; C594-C620 | HMMER_PFAM |
| | | | | | Fibrinogen beta and gamma chains, C-terminal globular domain: F1889-S2098 | HMMER_PFAM |
| | | | | | Fibronectin type III domain: V623-S701; L712-L795; L803-S882; L893-S974; L985-S1062; A1074-S1157; T1165-S1243; V1256-T1335; L1347-S1430; E1438-S1514; E1529-S1608; M1619-S1697; L1708-S1785; L1796-S1873 | HMMER_PFAM |
| | | | | | Cytosolic domain: M1-Q6 Transmembrane domain: L7-L25 Non-cytosolic domain: K26-A2110 | TMHMMER |
| | | | | | | |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------|-----------------------|---------------------|---------------------------------|-------------------------------|--|----------------------------------|
| 37 cont | | | | | Fibrinogen beta and gamma chains C-terminal domain proteins BL00514: V1921-G1957, E1962-T1974, Y2008-S2022, N2044-S2073, S2073-P2097 | BLIMPS_BLOCKS |
| | | | | | Fibrinogen beta and gamma chains C-terminal domain signature: D2035-E2085 | PROFILES CAN |
| | | | | | Type III EGF-like signature PR00011: G540-C558 | BLIMPS_PRINTS |
| | | | | | PRECURSOR GLYCOPROTEIN SIGNAL | BLAST_PROD OM |
| | | | | | FIBRINOGEN BLOOD COAGULATION CHAIN | |
| | | | | | PLASMA PROTEIN PLATELET PD001241: T306-I314; S1873-P2097 | |
| | | | | | GLYCOPROTEIN PRECURSOR TENASCIN SIGNAL MATRIX CYTOTACTIN ANTIGEN CELL TENASCIN-X EGF-LIKE PD004440: M4-R137 | BLAST_PROD OM |
| | | | | | PROTEIN TRANSCRIPTIONAL REPEAT | BLAST_PROD OM |
| | | | | | TRANSCRIPTION REGULATION DNA-BINDING NUCLEAR SHUTTLE CRAFT PUTATIVE PD014613: C161-P625 | |
| | | | | | GLYCOPROTEIN TENASCIN TENASCIN-X ANTIGEN PRECURSOR MATRIX CELL X TN HEXABRACHION PD000928: L1444-T1526; L1535-A1618 | BLAST_PROD OM |
| | | | | | FIBRINOGEN BETA/GAMMA DM00531 P24821 1946-2187: P1856-N2099 | BLAST_DOM O |
| | | | | | FIBRINOGEN BETA/GAMMA DM00531 S19694 1492-1734: P1856-N2099 | BLAST_DOM O |
| | | | | | FIBRINOGEN BETA/GAMMA DM00531 P10039 1554-1796: P1856-N2099 | BLAST_DOM O |

Table 3

| SEQ ID NO: | Incyte Polypeptide ID | Amino Acid Residues | Potential Phosphorylation Sites | Potential Glycosylation Sites | Signature Sequences, Domains and Motifs | Analytical Methods and Databases |
|------------------|-----------------------------|------------------------|---------------------------------------|----------------------------------|--|-------------------------------------|
| 37 | | | | | FIBRINOGEN BETA/GAMMA | BLAST_DOMO |
| cont | | | | | DM00531 JH0675 I096-1338: S1857-N2099 | |
| | | | | | Cell attachment sequence (RGD): R877-D879 | MOTIFS |
| | | | | | EGF-like domain signature 1: C174-C185, C205-C216, C236-C247, C268-C279, C299-C310, C330-C341, C361-C372, C392-C403, C423-C434, C454-C465, C485-C496, C516-C527, C547-C558, C578-C589, C609-C620 | MOTIFS |
| | | | | | EGF-like domain signature 2: C174-C185, C205-C216, C236-C247, C268-C279, C299-C310, C330-C341, C361-C372, C392-C403, C423-C434, C454-C465, C485-C496, C516-C527, C547-C558, C578-C589, C609-C620 | MOTIFS |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 38/7504868CB1/ 1469 | 1-729, 78-471, 85-669, 85-705, 85-1402, 207-436, 433-673, 544-625, 545-625, 562-1345, 585-1342, 601-1345, 622-1345, 626-1345, 683-1345, 684-1062, 688-745, 728-1068, 745-1399, 745-1400, 745-1461, 747-1332, 797-1418, 882-1234, 894-1469, 910-1465, 950-1076, 988-1126, 1019-1126, 1021-1126, 1022-1141 |
| 39/7504930CB1/ 5543 | 1-5460, 17-504, 38-505, 218-738, 220-620, 220-622, 220-737, 220-772, 220-819, 220-881, 220-905, 223-769, 230-597, 230-621, 230-702, 232-763, 282-591, 282-701, 282-702, 282-744, 286-732, 286-733, 312-861, 320-880, 321-438, 459-992, 615-1171, 615-1206, 660-1267, 675-996, 687-1214, 742-1022, 1062-1323, 1175-1307, 1197-1704, 1218-1834, 1227-1834, 1259-1919, 1337-1822, 1347-1629, 1442-2121, 1443-1914, 1454-1918, 1469-1914, 1497-1634, 1513-1914, 1519-2246, 1556-2105, 1610-2160, 1674-1786, 1742-2367, 1784-2118, 1784-2317, 1818-2271, 1863-1982, 1890-2457, 1899-2493, 1914-2569, 1946-2604, 2053-2400, 2118-2612, 2247-2488, 2247-2729, 2457-2613, 2503-2640, 2532-3117, 2678-2968, 2687-3228, 2691-2926, 2754-3196, 2775-3081, 2785-3417, 2957-3231, 3111-3392, 3190-3695, 3269-3883, 3269-3884, 3278-3938, 3408-4074, 3453-3682, 3465-4032, 3480-3773, 3481-4003, 3495-3976, 3535-4136, 3600-4166, 3609-4174, 3617-3857, 3645-4207, 3668-3878, 3668-4096, 3670-4017, 3679-4082, 3703-4343, 3708-4229, 3715-4234, 3762-4011, 3773-4251, 3786-4045, 3787-4054, 3844-3968, 3861-4078, 3878-4382, 3919-4532, 3936-4175, 3987-4643, 3994-4486, 3996-4190, 4006-4602, 4022-4235, 4022-4552, 4042-4406, 4042-4430, 4048-4705, 4054-4520, 4055-4361, |
| 39 cont | 4062-4519, 4070-4408, 4077-4485, 4078-4667, 4137-4677, 4139-4670, 4184-4409, 4184-4420, 4184-4467, 4184-4697, 4184-4747, 4184-4817, 4193-4504, 4208-4798, 4218-4457, 4234-4734, 4234-4877, 4241-4867, 4244-4496, 4245-4489, 4263-4480, 4269-5046, 4270-4822, 4295-4855, 4303-4835, 4319-4817, 4319-4888, 4325-4794, 4347-4624, 4414-4869, 4415-4698, 4418-4759, 4418-4763, 4421-4941, 4433-4654, 4434-5141, 4495-5014, 4501-5085, 4579-5177, 4588-5139, 4591-5103, 4607-5142, 4640-4789, 4640-5116, 4641-5140, 4661-5154, 4677-4799, 4711-5135, 4725-5005, 4730-5202, 4735-4959, 4735-5196, 4735-5202, 4736-4954, 4739-5154, 4741-5202, 4743-4970, 4743-5204, 4745-5148, 4760-5206, 4768-5202, 4774-5065, 4776-5153, 4779-5021, 4800-5081, 4801-5200, 4807-5206, 4824-5089, 4833-5147, 4833-5154, 4843-5153, 4843-5206, 4852-4995, 4852-5186, 4852-5201, 4856-5152, 4856-5154, 4859-5046, 4870-5078, 4872-5127, 4874-5157, 4877-5138, 4896-5259, 4954-5208, 4955-5143, 4955-5206, 4980-5208, 4982-5153, 5010-5157, 5039-5202, 5100-5206, 5272-5543, 5274-5515 |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 40/6610456CB1/ 4293 | 1-543, 16-658, 25-313, 34-139, 46-802, 47-344, 47-625, 52-504, 52-505, 53-882, 53-939, 58-504, 59-420, 59-501, 66-368, 71-702, 74-860, 136-414, 152-876, 190-1028, 204-455, 297-952, 316-372, 318-699, 321-652, 411-652, 635-1237, 814-1290, 850-1493, 863-1247, 956-1552, 961-1501, 1115-1381, 1120-1868, 1129-1984, 1298-1513, 1298-1537, 1298-1564, 1298-1792, 1316-1839, 1408-1476, 1413-1840, 1514-2091, 1518-2112, 1554-1788, 1565-1889, 1718-2161, 1738-2195, 1768-1955, 1775-2022, 1805-2006, 1881-2347, 1883-2160, 1925-2705, 1930-2188, 1935-2346, 1966-2459, 1974-2626, 1975-2281, 1991-2532, 2015-2307, 2036-2290, 2080-2700, 2110-2410, 2134-2720, 2194-2461, 2203-2478, 2220-2499, 2319-2681, 2352-2611, 2357-2645, 2367-2810, 2421-2589, 2454-2686, 2456-2936, 2570-3032, 2592-2932, 2683-2941, 2734-3496, 2743-3381, 2746-2890, 2766-3027, 2783-3406, 2821-3522, 2831-3466, 2883-3143, 2885-3479, 2901-3531, 2911-3519, 2919-3557, 2932-3215, 2933-3189, 2959-3456, 2964-3221, 2964-3518, 2964-3555, 3001-3534, 3009-3121, 3009-3233, 3061-3536, 3072-3331, 3080-3530, 3081-3536, 3087-3557, 3091-3559, 3098-3535, 3110-3536, 3116-3550, 3137-3542, 3138-3530, 3139-3542, 3146-3536, 3168-3534, 3175-3542, 3177-3536, 3183-3472, 3183-3536, 3216-3536, |
| 40 cont | 3237-3536, 3242-3510, 3243-3537, 3296-3522, 3296-3530, 3313-3536, 3324-3552, 3392-3531, 3454-3530, 3454-4293 |
| 41/7503573CB1/ 4777 | 1-4692, 395-855, 568-753, 569-815, 570-1122, 599-1179, 607-1221, 643-930, 703-993, 723-1293, 766-1339, 790-1188, 812-1475, 829-1365, 842-1179, 845-1194, 867-1156, 879-1439, 892-1211, 901-1490, 911-1197, 922-1504, 939-1500, 950-975, 953-1224, 1004-1111, 1028-1563, 1031-1499, 1056-1340, 1099-1406, 1130-1528, 1191-1806, 1199-1426, 1236-1459, 1242-1845, 1249-1859, 1250-1487, 1254-1769, 1269-1550, 1448-2112, 1456-2168, 1477-1766, 1477-1935, 1481-1691, 1488-2054, 1493-1784, 1499-1788, 1499-2092, 1500-2126, 1514-2302, 1526-1825, 1526-2001, 1535-1801, 1539-2169, 1553-2092, 1573-2107, 1575-2151, 1589-2162, 1604-2194, 1604-2233, 1641-1719, 1641-1731, 1648-1731, 1650-1731, 1653-1731, 1664-1947, 1689-1731, 1692-2157, 1704-1897, 1720-1949, 1730-2266, 1743-2053, 1743-2243, 1767-1857, 1774-1857, 1779-1869, 1779-2037, 1791-1857, 1813-2128, 1818-1857, 1847-2120, 1869-2242, 1913-2147, 1949-2162, 1963-2251, 2004-2283, 2017-2102, 2030-2142, 2040-2276, 2107-2254, 2131-2285, 2143-2228, 2275-2528, 2278-2576, 2280-2516, 2280-2543, 2280-2885, 2282-2482, 2282-2838, 2284-2771, 2290-2738, 2290-2743, 2303-2516, 2317-2497, 2344-2762, 2344-2909, 2345-2603, 2346-2612, 2346-2628, 2354-2644, 2357-2618, 2373-2598, 2373-2857, 2384-2838, 2403-2590, 2420-2787, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 41 cont | 2446-2474, 2446-2475, 2446-2476, 2450-3156, 2462-2780, 2468-2746, 2472-2728, 2472-2765, 2490-2749, 2508-2702, 2520-2861, 2525-2806, 2532-2791, 2532-2801, 2538-2758, 2541-2869, 2543-2874, 2552-2814, 2552-2843, 2553-2793, 2562-2811, 2562-2856, 2563-2728, 2586-3224, 2593-2921, 2600-2849, 2615-2907, 2617-2900, 2644-2918, 2648-3290, 2651-2932, 2658-2896, 2658-2921, 2658-3164, 2660-2962, 2674-2869, 2677-3190, 2686-3298, 2690-2959, 2692-2885, 2697-3291, 2700-2945, 2701-3210, 2706-3278, 2707-3232, 2712-2952, 2712-3015, 2733-3307, 2739-3023, 2741-3034, 2742-3245, 2781-3044, 2785-3292, 2794-3093, 2794-3262, 2815-2849, 2817-3108, 2824-3100, 2829-3138, 2834-3094, 2855-3131, 2861-2999, 2861-3098, 2861-3139, 2865-3187, 2870-3212, 2880-3106, 2880-3111, 2900-3193, 2916-3199, 2917-3207, 2934-3199, 2947-2981, 2949-2981, 2951-3448, 2954-3222, 2972-3261, 2972-3440, 2976-3245, 2986-3238, 2998-3175, 3001-3304, 3022-3250, 3030-3470, 3040-3266, 3047-3307, 3050-3202, 3071-3307, 3091-3258, 3122-3302, 3122-3307, 3159-3365, 3219-3465, 3268-3695, 3290-3887, 3307-3958, 3317-3437, 3330-3763, 3347-3725, 3372-3907, 3372-3958, 3374-3773, 3375-3669, 3391-4078, 3393-3958, 3396-3973, 3420-3903, |
| 41 cont | 3428-3927, 3430-3927, 3435-3669, 3476-3716, 3480-3936, 3480-4034, 3481-3772, 3482-3737, 3482-3806, 3482-3875, 3482-3910, 3482-4093, 3483-3747, 3483-3829, 3483-3876, 3484-3960, 3485-3695, 3485-3781, 3485-3848, 3485-3877, 3485-3898, 3485-3900, 3485-3904, 3485-3921, 3485-3945, 3485-3947, 3485-3959, 3485-3994, 3485-4001, 3485-4015, 3485-4021, 3485-4027, 3485-4029, 3485-4065, 3485-4093, 3485-4108, 3485-4110, 3485-4133, 3490-3958, 3490-3983, 3490-4068, 3494-3901, 3494-3909, 3501-3880, 3503-3800, 3506-3680, 3509-3761, 3509-3827, 3509-3836, 3509-3859, 3509-3878, 3509-3907, 3509-3931, 3509-3961, 3509-3962, 3509-3966, 3509-3969, 3509-3979, 3512-3677, 3515-3918, 3517-4070, 3518-4163, 3519-4080, 3519-4160, 3519-4306, 3527-3699, 3527-3780, 3548-3788, 3549-3791, 3550-3805, 3557-3837, 3569-3814, 3574-3743, 3575-3787, 3585-3746, 3585-3831, 3585-3838, 3590-4140, 3592-3851, 3594-3819, 3597-4189, 3603-3763, 3603-3900, 3614-3900, 3614-4097, 3615-3802, 3615-4262, 3616-4074, 3622-4032, 3627-3908, 3630-3996, 3636-3724, 3644-4114, 3647-3913, 3648-3906, 3664-3903, 3668-3866, 3668-3971, 3670-4297, 3677-3919, 3677-3928, 3693-3870, 3693-3919, 3693-4265, 3694-4329, |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 41 cont | 3701-3972, 3722-4036, 3724-4184, 3729-3944, 3733-3971, 3734-3928, 3736-3982, 3743-3765, 3749-4133, 3753-3979, 3755-4013, 3755-4375, 3776-3944, 3779-4028, 3780-4421, 3782-4010, 3788-4010, 3789-4073, 3790-4059, 3793-4024, 3797-4057, 3801-3913, 3805-4266, 3808-4416, 3841-3971, 3842-4053, 3854-4484, 3857-4154, 3868-3946, 3877-4099, 3877-4135, 3877-4442, 3879-4123, 3892-4185, 3893-4687, 3900-4180, 3900-4376, 3900-4452, 3902-4165, 3903-4161, 3903-4181, 3904-4564, 3907-4012, 3910-4174, 3928-4201, 3930-4153, 3936-4618, 3938-4633, 3951-4635, 3955-4222, 3956-4642, 3959-4297, 3961-4692, 3966-4691, 3981-4172, 3985-4256, 3986-4384, 3989-4566, 3989-4654, 4005-4265, 4011-4298, 4017-4260, 4018-4144, 4018-4223, 4020-4283, 4020-4286, 4027-4213, 4030-4300, 4038-4334, 4053-4298, 4055-4640, 4060-4648, 4064-4261, 4075-4293, 4079-4302, 4084-4384, 4089-4338, 4089-4378, 4094-4380, 4094-4527, 4096-4370, 4104-4383, 4105-4226, 4112-4389, 4114-4361, 4114-4401, 4120-4386, 4125-4696, 4129-4412, 4132-4412, 4133-4446, 4149-4671, 4150-4392, 4160-4551, 4171-4709, 4190-4450, 4204-4441, 4211-4593, 4212-4692, 4227-4691, 4231-4683, 4236-4690, 4243-4693, 4247-4692, 4250-4494, 4257-4692, 4285-4441, 4295-4692, |
| 41 cont | 4297-4692, 4299-4694, 4300-4697, 4326-4692, 4333-4362, 4345-4694, 4345-4695, 4350-4694, 4355-4694, 4357-4679, 4360-4696, 4364-4661, 4389-4602, 4389-4694, 4438-4681, 4438-4692, 4444-4657, 4451-4659, 4454-4692, 4463-4691, 4484-4702, 4490-4647, 4494-4706, 4500-4692, 4504-4707, 4512-4682, 4513-4696, 4517-4718, 4522-4707, 4523-4715, 4544-4777, 4546-4767, 4548-4692, 4556-4731, 4562-4692 |
| 42/7505057CB1/ 1463 | 1-197, 1-1086, 28-138, 43-146, 57-284, 57-353, 57-724, 58-230, 59-182, 63-805, 235-509, 235-732, 235-830, 235-845, 235-887, 235-888, 236-413, 236-470, 236-476, 236-496, 236-509, 236-627, 236-820, 236-841, 236-906, 237-741, 237-829, 239-750, 240-630, 240-885, 241-414, 241-833, 242-511, 245-498, 245-889, 247-543, 247-749, 252-479, 252-504, 252-520, 252-528, 252-534, 252-765, 253-456, 255-876, 258-565, 261-533, 261-540, 261-557, 262-888, 263-523, 264-676, 264-737, 265-888, 266-841, 267-531, 269-533, 270-830, 271-379, 271-537, 271-551, 271-975, 276-555, 276-870, 278-404, 278-504, 278-525, 278-801, 283-538, 284-528, 284-529, 285-560, 285-566, 285-888, 289-559, 289-900, 292-548, 293-600, 294-527, 294-553, 296-494, 296-561, 296-566, 296-602, 296-605, 297-584, 297-824, 298-909, 298-960, 300-558, 300-584, 301-562, 302-570, 309-540, 309-578, 309-806, 310-577, 311-443, 312-587, 313-561, 315-619, 317-474, 317-630, 317-943, 317-957, 318-677, 318-779, 319-558, 319-580, 319-770, 321-637, 321-951, 322-959, 324-535, 326-946, 330-830, 334-917, 336-612, 336-619, 337-551, 337-815, 338-640, 342-542, 346-888, 347-536, 350-834, 351-641, 352-888, 354-638, 355-650, 356-597, 357-623, |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 42 cont | 358-620, 358-631, 362-615, 363-492, 364-598, 364-637, 365-627, 365-1041, 368-693, 370-957, 374-653, 374-667, 375-931, 375-1057, 376-871, 376-901, 376-965, 378-602, 379-627, 381-588, 381-1014, 382-838, 387-673, 389-639, 389-642, 389-662, 391-649, 391-671, 391-711, 393-890, 395-670, 395-681, 397-783, 398-1009, 399-664, 400-610, 400-647, 402-504, 403-666, 403-678, 403-684, 403-866, 406-902, 407-616, 407-726, 407-772, 409-659, 409-698, 409-707, 411-664, 413-583, 413-622, 413-630, 413-635, 413-641, 413-684, 415-686, 415-999, 416-605, 416-672, 418-1015, 421-661, 421-671, 422-625, 422-697, 422-977, 423-694, 424-661, 424-691, 424-958, 425-723, 428-732, 434-610, 434-730, 443-635, 443-653, 443-719, 449-696, 450-690, 451-575, 451-696, 451-714, 451-721, 451-890, 454-626, 454-738, 456-721, 461-677, 461-720, 467-884, 467-987, 468-999, 469-975, 471-844, 472-719, 472-731, 472-759, 474-746, 475-729, 476-884, 483-641, 489-734, 489-1133, 490-955, 493-742, 493-772, 493-777, 494-792, 495-767, 496-742, 496-750, 496-758, 496-826, 496-1062, 496-1111, 497-626, 497-761, 498-766, 499-734, 499-746, 499-750, 499-769, 499-772, 499-780, 500-614, 500-749, 500-755, 500-772, 500-791, |
| 42 cont | 501-962, 502-765, 504-775, 505-811, 506-757, 506-759, 506-822, |
| 42 cont | 507-768, 508-739, 508-740, 510-1169, 511-811, 511-964, 514-702, |
| 42 cont | 514-940, 518-653, 520-804, 520-822, 525-781, 526-742, 527-784, 528-957, 529-777, 536-672, 537-770, 537-819, 540-767, 544-801, 546-795, 547-1064, 549-796, 551-827, 553-964, 556-778, 556-807, 557-814, 558-717, 558-770, 559-816, 559-833, 559-890, 559-1029, 559-1060, 560-799, 560-1159, 567-777, 567-816, 568-803, 568-814, 572-823, 572-838, 575-883, 575-1025, 577-816, 579-798, 580-783, 580-1015, 585-829, 585-843, 586-839, 590-828, 591-845, 595-893, 596-1100, 599-848, 599-862, 599-873, 600-814, 601-874, 605-757, 606-894, 609-876, 611-1003, 612-840, 612-847, 612-855, 612-870, 612-874, 612-876, 612-894, 613-842, 613-856, 613-870, 613-899, 615-883, 618-964, 619-804, 620-877, 620-879, 620-906, 620-908, 620-925, 620-929, 621-912, 624-852, 626-871, 627-921, 627-972, 630-829, 631-929, 634-898, 636-917, 638-827, 638-871, 638-875, 638-909, 638-1004, 640-872, 641-805, 641-880, 642-969, 643-971, 646-889, 646-1083, 647-916, 647-922, 647-927, 647-941, 647-971, 648-898, 651-949, 651-963, 651-1033, 652-946, 652-957, 653-865, 653-916, 659-940, 665-919, 667-911, 671-948, 671-1029, 672-905, 677-930, 680-938, 682-962, 684-921, 688-950, 688-1171, 690-783, 697-962, 698-970, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 42 cont | 712-994, 717-978, 718-1019, 719-933, 719-991, 719-995, 721-960, 731-944, 731-1000, 733-966, 733-988, 733-1012, 733-1059, 735-1007, 737-975, 737-982, 737-1007, 737-1072, 737-1203, 741-992, 741-1014, 744-1015, 745-1005, 749-1010, 751-1010, 753-982, 753-1017, 760-1034, 760-1059, 762-1037, 762-1040, 762-1051, 762-1071, 763-1045, 764-975, 764-1082, 765-1042, 766-1070, 771-994, 771-1063, 772-1034, 772-1060, 772-1101, 775-892, 775-1016, 775-1027, 775-1034, 775-1044, 775-1053, 776-1071, 778-1463, 779-1051, 779-1096, 780-1018, 782-940, 782-1025, 783-1032, 785-1055, 790-975, 791-1080, 793-1046, 795-1058, 802-1086, 802-1109, 809-1023, 809-1063, 809-1080, 809-1088, 809-1090, 810-1095, 812-1080, 813-1018, 814-1082, 816-985, 818-1079, 820-978, 820-1080, 820-1147, 821-1074, 821-1075, 823-1068, 833-1058, 833-1086, 833-1109, 835-1078, 839-1096, 847-1058, 849-1109, 850-1097, 853-1075, 866-981, 874-1077, 874-1085, 882-1118, 883-1120, 887-1032, 889-1144, 903-1083, 903-1089, 906-1071, 936-1133, 939-1015, 957-1105, 972-994 |
| 43/90116002CBI/ 1259 | 1-911, 97-1058, 150-1078, 172-1058, 178-1057, 182-1057, 202-1057, 216-1077, 229-1078, 249-1057, 290-1077, 296-1077, 345-1057, 362-1077, 366-1070, 404-1071, 410-1057, 449-1142, 500-1259, 567-1223, 578-1057, 578-1100, 644-1259, 652-1154, 665-1238, 720-1259, 722-1259, 728-1259, 789-1259 |
| 44/039283CBI/ 3548 | 1-539, 272-1017, 274-1022, 284-932, 315-1006, 318-1043, 328-777, 344-950, 369-1041, 458-640, 487-999, 490-791, 495-933, 498-718, 498-833, 507-770, 521-1140, 526-1056, 540-816, 575-1064, 653-1200, 767-1289, 812-1394, 861-1171, 883-1495, 942-1178, 993-1171, 1032-1522, 1072-1542, 1075-1533, 1159-1532, 1208-1540, 1250-1540, 1286-1540, 1309-1540, 1387-1533, 1400-1540, 1416-1540, 1431-1540, 1440-1540, 1462-1993, 1478-1540, 1488-1540, 1528-1658, 1528-1763, 1528-1864, 1528-1891, 1528-1901, 1528-1910, 1577-1926, 1619-2230, 1655-2269, 1678-2278, 1756-2442, 1769-2076, 1785-2349, 1800-2399, 1809-2242, 1831-1910, 1847-2434, 1857-2419, 1861-2550, 1882-2494, 1883-2148, 1918-2459, 1927-2174, 1938-2527, 1947-2550, 1950-2220, 1955-2152, 1955-2451, 1963-2292, 1984-2129, 1987-2360, 1990-2217, 1995-2634, 1998-2554, 2015-2523, 2026-2284, 2034-2327, 2040-2517, 2047-2547, 2047-2600, 2047-2610, 2047-2621, 2062-2254, 2062-2334, 2069-2568, 2080-2739, 2107-2671, 2107-2741, 2129-2685, 2130-2741, 2134-2707, 2135-2430, 2137-2767, 2152-2721, 2156-2767, 2180-2860, 2182-2848, 2185-2847, 2192-2840, 2202-2505, 2204-2478, 2216-2861, 2218-2701, 2218-2776, 2227-2674, 2239-2481, 2243-2838, 2245-2892, 2246-2856, 2260-2845, 2261-2814, 2262-2886, |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 44 cont | 2265-2787, 2268-2884, 2272-2895, 2273-2549, 2277-2523, 2291-2895, 2294-2753, 2319-2564, 2337-2610, 2337-2820, 2353-2862, 2354-2592, 2354-2597, 2356-2886, 2364-2717, 2366-2942, 2370-2670, 2374-2920, 2375-2909, 2382-2650, 2393-2723, 2409-3013, 2411-3034, 2426-2706, 2439-2768, 2439-2891, 2451-2888, 2458-2711, 2458-2721, 2476-2732, 2476-3030, 2490-3182, 2505-2784, 2507-2754, 2514-3027, 2516-2798, 2516-2804, 2518-2831, 2522-2999, 2537-3059, 2540-2938, 2566-3191, 2572-3042, 2572-3189, 2572-3297, 2576-3221, 2590-2779, 2590-2792, 2590-3082, 2605-2868, 2607-2850, 2643-2988, 2662-2918, 2680-2902, 2762-3025, 2769-3025, 2852-3528, 2879-3135, 2919-3525, 2919-3533, 2934-3222, 2939-3533, 2950-3538, 2962-3225, 2990-3535, 3008-3529, 3035-3294, 3068-3335, 3068-3533, 3068-3548, 3145-3532, 3155-3307, 3155-3357, 3169-3475, 3227-3427, 3227-3499 |
| 45/7505082CBI/ 776 | 1-116, 1-776, 29-152, 109-696, 161-583, 161-586, 161-702, 161-718, 264-566, 327-768, 329-508, 331-542, 348-569, 353-450, 369-639, 384-562, 456-727, 466-683, 474-599, 478-743, 546-739, 546-775 |
| 46/7505139CBI/ 2521 | 1-268, 1-2441, 52-361, 53-282, 56-325, 65-619, 68-322, 70-536, 74-306, 74-337, 75-215, 75-361, 76-321, 76-342, 76-368, 77-333, 77-340, 77-341, 77-425, 77-465, 78-251, 78-343, 78-345, 79-523, 80-233, 80-330, 80-355, 80-523, 81-352, 83-355, 83-375, 86-305, 86-306, 86-370, 86-493, 86-523, 87-379, 87-493, 89-594, 90-322, 92-197, 92-364, 93-364, 95-338, 95-353, 95-361, 95-363, 95-392, 95-432, 96-313, 96-372, 96-387, 96-464, 97-181, 97-331, 97-336, 97-363, 97-376, 97-385, 97-387, 97-523, 98-362, 98-607, 99-310, 103-327, 103-352, 105-267, 105-348, 105-371, 105-619, 108-344, 112-354, 116-339, 116-370, 117-523, 119-364, 127-511, 130-475, 130-523, 130-618, 131-409, 131-512, 135-362, 136-520, 137-362, 137-401, 137-420, 138-336, 138-421, 139-387, 139-444, 140-329, 140-349, 140-415, 140-436, 141-386, 150-393, 150-406, 150-416, 150-419, 152-442, 156-419, 156-423, 157-454, 185-353, 185-453, 185-484, 185-523, 210-493, 218-520, 218-538, 219-523, 238-523, 245-511, 433-523, 619-913, 630-1020, 630-1021, 631-714, 631-792, 637-1128, 643-945, 649-923, 650-841, 658-899, 663-893, 663-948, 665-1141, 669-919, 669-928, 669-1109, 670-929, 670-966, 670-1258, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 46 cont | 675-949, 680-1233, 688-902, 689-953, 690-881, 690-1061, 693-1247, 702-908, 707-979, 709-986, 713-893, 714-977, 717-1609, 718-929, 718-999, 718-1010, 723-1033, 726-1455, 728-879, 728-998, 730-1024, 735-1251, 735-1286, 737-984, 737-987, 737-1227, 737-1232, 737-1274, 738-983, 740-1642, 743-1609, 744-989, 754-981, 754-1034, 755-1237, 760-972, 762-998, 765-1034, 767-1039, 786-1607, 787-1609, 790-1038, 791-976, 791-996, 798-1065, 804-1609, 807-1106, 809-1027, 810-1066, 810-1283, 814-1050, 814-1053, 816-1091, 816-1093, 818-1108, 821-1107, 824-1204, 827-1010, 827-1144, 827-1286, 832-1086, 834-1435, 836-1593, 837-963, 838-1531, 839-1100, 839-1120, 840-1108, 842-1203, 846-1609, 847-1137, 850-1609, 854-1119, 860-1122, 861-1114, 875-1146, 880-1086, 886-1125, 888-1677, 893-1171, 900-1609, 908-1174, 909-1225, 913-1078, 914-1155, 914-1565, 915-1244, 917-1161, 917-1170, 917-1191, 919-1219, 919-1326, 923-1414, 930-1542, 932-1168, 932-1492, 939-1392, 940-1204, 940-1238, 942-1187, 950-1214, 951-1209, 953-1228, 957-1232, 958-1210, 959-1176, 960-1242, 960-1427, 961-1192, 961-1219, 964-1231, 965-1086, 974-1609, 975-1200, 977-1233, 982-1450, 983-1082, 988-1226, |
| 46 cont | 988-1266, 994-1176, 1006-1169, 1006-1262, 1006-1305, 1008-1283, 1016-1272, 1016-1654, 1035-1279, 1035-1308, |
| 46 cont | 1035-1514, 1035-1828, 1038-1336, 1038-1352, 1039-1313, 1042-1623, 1044-1277, 1044-1654, 1045-1291, 1054-1623, 1055-1290, 1058-1306, 1059-1279, 1068-1339, 1076-1211, 1078-1366, 1084-1383, 1086-1673, 1087-1262, 1087-1340, 1087-1362, 1093-1277, 1095-1308, 1095-1321, 1095-1346, 1103-1370, 1103-1664, 1106-1486, 1106-1583, 1111-1369, 1121-1398, 1124-1368, 1128-1367, 1132-1587, 1132-1813, 1133-1178, 1135-1172, 1135-1173, 1135-1174, 1135-1386, 1136-1178, 1138-1178, 1145-1357, 1156-1403, 1156-1405, 1156-1416, 1156-1429, 1156-1557, 1164-1424, 1170-1403, 1170-1435, 1173-1666, 1174-1412, 1175-1209, 1175-1210, 1175-1211, 1175-1212, 1175-1216, 1175-1443, 1175-1607, 1175-1849, 1178-1216, 1178-1820, 1186-1457, 1193-1438, 1193-1476, 1193-1485, 1194-1440, 1194-1447, 1194-1470, 1195-1666, 1207-1462, 1207-1864, 1208-1446, 1208-1450, 1208-1462, 1208-1511, 1213-1664, 1215-1528, 1223-1652, 1228-1664, 1229-1479, 1229-1537, 1232-1529, 1232-1530, 1234-1472, 1235-1503, 1237-1428, 1237-1492, 1237-1505, 1241-1666, 1242-1433, 1242-1664, 1243-1441, 1243-1526, 1244-1488, 1244-1512, 1244-1554, 1245-1663, 1245-1668, 1246-1632, 1251-1664, 1261-1661, 1262-1666. |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 46 cont | 1262-1801, 1263-1537, 1265-1476, 1267-1661, 1268-1527, 1268-1528, 1268-1587, 1269-1532, 1274-1666, 1275-1548, 1275-1666, 1278-1521, 1281-1530, 1281-1554, 1281-1565, 1282-1430, 1282-1544, 1293-1445, 1294-1661, 1298-1664, 1302-1526, 1310-1666, 1312-1667, 1315-1587, 1315-1612, 1316-1575, 1323-1566, 1323-1907, 1329-2008, 1332-1664, 1333-1600, 1333-1667, 1342-1666, 1344-1666, 1350-1579, 1351-1606, 1355-1602, 1355-1617, 1359-1666, 1362-2084, 1367-1627, 1371-1656, 1371-1664, 1373-1544, 1373-1616, 1380-1639, 1383-1621, 1383-1687, 1388-1656, 1389-1606, 1389-1654, 1390-1692, 1397-1658, 1398-1676, 1402-1661, 1405-1662, 1407-1663, 1414-1642, 1414-1666, 1414-1688, 1418-1820, 1419-1735, 1426-1661, 1427-1886, 1435-2070, 1438-1671, 1442-1729, 1444-1882, 1452-1755, 1456-1687, 1462-1672, 1462-1915, 1473-1720, 1473-1731, 1475-1724, 1475-1764, 1476-1723, 1476-1758, 1477-1664, 1482-1761, 1486-1666, 1486-1768, 1490-1723, 1491-1775, 1493-1740, 1493-1741, 1496-1928, 1500-1914, 1502-1705, 1502-1721, 1506-1738, 1507-1744, 1509-1712, 1510-2357, 1513-1817, 1514-1994, 1515-1809, 1517-1607, 1519-1766, 1519-1806, 1529-1745, 1531-1697, 1532-2183, 1538-1804, 1543-1667, 1544-1808, 1545-1792, 1546-1821, 1547-1781, 1548-1773, 1548-1794, 1548-2107, |
| 46 cont | 1551-1829, 1552-1770, |
| 46 cont | 1552-1779, 1553-1843, 1554-1829, 1555-1679, 1556-1913, 1556-2376, 1558-1806, 1558-1808, 1558-1826, 1558-1841, 1558-1845, 1558-2081, 1559-1851, 1561-1859, 1562-1784, 1562-1814, 1571-1851, 1572-1810, 1575-1862, 1576-1821, 1579-1824, 1579-1829, 1579-1841, 1587-1825, 1587-1836, 1591-1880, 1591-1887, 1594-1848, 1595-1868, 1595-1872, 1596-1872, 1598-1827, 1598-1847, 1601-2262, 1602-2198, 1603-1666, 1604-1850, 1605-1856, 1615-2065, 1617-1916, 1618-1870, 1618-1913, 1621-1921, 1622-1834, 1622-1867, 1622-1912, 1628-1912, 1630-1867, 1631-2385, 1632-2357, 1633-1874, 1636-2362, 1646-2354, 1649-2205, 1650-1877, 1652-1865, 1652-1900, 1652-1922, 1653-1875, 1655-1869, 1655-1885, 1656-1940, 1657-2426, 1659-1755, 1659-1875, 1663-1849, 1663-1852, 1663-1929, 1663-1930, 1663-1940, 1663-1953, 1663-2060, 1663-2138, 1664-1894, 1664-1994, 1664-2111, 1666-1821, 1668-1805, 1672-1933, 1675-1941, 1683-2118, 1689-1950, 1691-1942, 1692-1953, 1699-1967, 1699-1979, 1699-1998, 1699-2365, 1701-1973, 1702-1879, 1702-1931, 1702-2377, 1704-1875, 1706-1976, 1712-1950, 1712-1971, 1715-2357, 1715-2410, 1718-1855, 1719-1954, 1722-2010, 1722-2048, 1723-1916, 1731-2426, 1732-1996, 1732-2019, 1735-2428, 1738-1980, 1738-2012, 1738-2027, 1739-1952, 1739-2357, 1740-2363, 1740-2401, |
| 46 cont | 1743-1990, 1744-2419, |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 46 cont | 1745-1922, 1748-1952, 1748-1954, 1750-1975, 1751-2033, 1757-1959, 1757-2022, 1757-2075, 1758-1864, 1761-2139, 1765-2326, 1767-2385, 1768-1957, 1768-1984, 1768-1999, 1768-2026, 1768-2053, 1770-1967, 1770-1971, 1772-2015, 1774-2357, 1775-2030, 1776-2005, 1776-2013, 1776-2042, 1776-2077, 1781-2032, 1783-1963, 1783-2043, 1783-2063, 1786-1942, 1786-2053, 1786-2056, 1786-2085, 1787-2011, 1791-2426, 1793-2292, 1793-2306, 1793-2420, 1794-2013, 1794-2091, 1795-2076, 1795-2087, 1797-2318, 1799-2041, 1801-2240, 1803-2043, 1810-2033, 1810-2047, 1810-2059, 1810-2060, 1810-2074, 1816-2146, 1817-2414, 1824-2058, 1824-2074, 1825-1969, 1826-1939, 1826-1960, 1826-2042, 1826-2109, 1826-2302, 1830-2067, 1831-2121, 1831-2382, 1831-2444, 1834-2076, 1834-2083, 1834-2094, 1834-2171, 1835-2100, 1835-2371, 1835-2380, 1835-2388, 1836-2459, 1837-2262, 1837-2264, 1838-1966, 1838-2120, 1838-2121, 1839-2243, 1843-2074, 1844-2120, 1846-2078, 1847-2078, 1847-2085, 1847-2423, 1848-2091, 1849-2046, 1850-2355, 1855-2322, 1856-2112, 1857-2253, 1861-2140, 1862-2101, 1862-2163, 1863-2447, 1864-2101, 1865-2353, 1865-2358, 1866-2126, 1867-2105, 1868-2141, 1868-2418, 1869-2392, 1870-2202, 1871-1980, 1871-2137, |
| 46 cont | 1871-2361, 1875-2125, 1875-2149, 1886-1995, |
| 46 cont | 1887-2133, 1894-2155, 1895-2149, 1895-2158, 1896-2118, 1897-2148, 1898-2452, 1901-2104, 1903-2194, 1904-2316, 1906-2453, 1911-2337, 1912-2161, 1912-2203, 1917-2091, 1918-2217, 1922-2169, 1922-2204, 1928-2187, 1928-2385, 1929-2210, 1930-2357, 1931-2210, 1931-2433, 1936-2177, 1937-2165, 1937-2170, 1937-2453, 1938-2426, 1940-2217, 1945-2176, 1945-2440, 1947-2439, 1948-2281, 1948-2426, 1950-2433, 1953-2425, 1953-2435, 1954-2232, 1954-2429, 1956-2459, 1958-2221, 1960-2223, 1960-2426, 1963-2426, 1964-2121, 1966-2360, 1966-2426, 1967-2246, 1968-2171, 1968-2228, 1968-2428, 1969-2426, 1970-2428, 1972-2417, 1973-2421, 1973-2428, 1973-2432, 1975-2214, 1975-2426, 1977-2423, 1977-2428, 1977-2435, 1977-2444, 1978-2426, 1981-2440, 1982-2157, 1983-2161, 1983-2183, 1983-2242, 1983-2416, 1985-2456, 1986-2207, 1986-2306, 1987-2232, 1987-2250, 1987-2446, 1988-2252, 1989-2428, 1991-2455, 1992-2433, 1994-2105, 1994-2427, 1994-2430, 1995-2423, 1995-2428, 1996-2423, 1997-2240, 1999-2448, 2000-2426, 2000-2450, 2001-2428, 2004-2426, 2005-2451, 2006-2426, 2006-2433, 2006-2453, 2008-2434, 2009-2426, 2010-2426, 2011-2428, 2012-2431, 2013-2426, 2013-2430, |
| 46 cont | 2013-2440, 2015-2408, 2016-2224, 2016-2229, 2016-2241, 2016-2305, 2016-2428, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 46 cont | 2016-2430, 2017-2306, 2017-2320, 2017-2428, 2017-2442, 2018-2303, 2018-2309, 2019-2272, 2019-2295, 2019-2428, 2020-2253, 2020-2258, 2020-2275, 2020-2430, 2021-2433, 2024-2431, 2025-2249, 2026-2430, 2027-2426, 2027-2428, 2028-2426, 2028-2431, 2029-2426, 2029-2430, 2030-2262, 2031-2198, 2032-2430, 2034-2430, 2035-2283, 2035-2308, 2035-2356, 2037-2348, 2037-2426, 2037-2430, 2038-2425, 2038-2426, 2040-2459, 2044-2278, 2045-2426, 2046-2426, 2047-2328, 2047-2426, 2052-2208, 2052-2272, 2052-2292, 2052-2426, 2052-2435, 2053-2430, 2065-2342, 2067-2349, 2067-2373, 2068-2333, 2069-2315, 2069-2445, 2070-2275, 2070-2276, 2070-2313, 2075-2428, 2077-2360, 2077-2370, 2081-2445, 2082-2430, 2083-2425, 2086-2426, 2086-2430, 2087-2431, 2088-2345, 2088-2426, 2088-2433, 2089-2324, 2089-2428, 2089-2432, 2090-2426, 2090-2428, 2090-2430, 2091-2426, 2092-2349, 2092-2430, 2092-2434, 2093-2345, 2095-2426, 2095-2430, 2102-2378, 2102-2428, 2104-2297, 2106-2324, 2107-2343, 2109-2349, 2109-2366, 2110-2426, 2112-2443, 2113-2400, 2115-2388, 2115-2415, 2115-2427, |
| 46 cont | 2118-2328, 2123-2424, 2125-2351, 2125-2439, 2136-2423, 2136-2426, 2139-2428, 2143-2430, 2143-2432, 2144-2367, 2145-2428, 2151-2430, |
| 46 cont | 2151-2437, 2153-2439, 2155-2428, 2156-2427, 2156-2428, 2159-2448, 2160-2426, 2160-2430, 2166-2415, 2167-2436, 2168-2426, 2176-2426, 2178-2406, 2181-2425, 2181-2435, 2185-2426, 2189-2445, 2190-2427, 2192-2433, 2193-2447, 2194-2432, 2195-2440, 2196-2426, 2196-2430, 2197-2413, 2198-2428, 2199-2426, 2205-2428, 2217-2434, 2218-2335, 2218-2447, 2231-2428, 2232-2412, 2234-2352, 2243-2454, 2243-2483, 2244-2444, 2248-2428, 2249-2430, 2263-2426, 2263-2469, 2268-2463, 2270-2430, 2283-2397, 2283-2430, 2297-2419, 2308-2432, 2312-2521, 2320-2465, 2321-2467, 2355-2458 |
| 47/7505234CB1/ 1884 | 1-301, 1-483, 1-584, 1-624, 1-724, 1-1884, 2-294, 2-523, 2-663, 4-624, 39-224, 39-537, 46-357, 47-563, 48-863, 51-851, 54-894, 58-651, 78-593, 222-771, 274-798, 311-1033, 329-752, 401-794, 433-557, 473-908, 548-1364, 550-1073, 550-1189, 557-1367, 607-1226, 620-1172, 774-1244, 782-1359, 800-1416, 805-1407, 806-1164, 806-1445, 808-1067, 809-1067, 1365-1835, 1372-1835, 1526-1830, 1528-1837, 1554-1831, 1620-1831 |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 48/7500227CB1/ 1132 | 1-284, 1-299, 2-250, 48-489, 48-493, 49-489, 323-420, 414-1069, 415-539, 420-698, 420-738, 428-1029, 435-718, 435-1067, 441-1059, 452-1103, 455-711, 461-1038, 465-905, 469-1068, 478-727, 498-748, 513-930, 513-952, 523-902, 529-1033, 534-1066, 536-910, 540-1087, 543-984, 551-728, 557-979, 581-787, 584-961, 587-884, 590-1127, 592-1081, 599-1042, 610-887, 614-1104, 617-1081, 619-1092, 625-921, 625-1082, 627-1082, 629-1080, 633-886, 635-1082, 641-910, 642-1064, 648-941, 650-1081, 650-1103, 653-1081, 654-1081, 657-1083, 660-925, 660-1103, 660-1109, 661-1083, 664-1084, 665-1084, 666-1084, 667-921, 670-1109, 671-948, 672-1081, 674-1082, 683-1088, 685-1085, 686-1096, 686-1122, 687-978, 687-1132, 688-1041, 688-1103, 700-1093, 702-1106, 703-1103, 709-897, 711-993, 717-1081, 720-1006, 731-1020, 731-1026, 740-912, 745-1103, 749-1108, 755-1095, 760-1065, 762-1083, 798-1132, 799-1082, 807-1127, 808-1103, 810-1104, 817-1103, 820-1082, 860-1132, 872-1092, 882-1081, 893-961, 904-1091, 904-1093, 917-1132, 928-1077, 928-1109, 953-1078, 955-1102, 974-1063, 981-1104, 981-1106, 990-1132, 999-1082, 1005-1081, 1012-1088 |
| 49/7503676CB1/ 2391 | 1-432, 4-423, 22-389, 24-457, 24-1995, 53-432, 63-570, 64-398, 115-432, 266-432, 436-1190, 485-542, 568-702, 568-831, 592-1210, 593-1145, 618-1165, 621-1202, 627-994, 635-1195, 638-1220, 649-1102, 659-1055, 667-1179, 673-1217, 706-1436, 724-1203, 738-1263, 738-1264, 749-1188, 761-1355, 764-1397, 764-1455, 771-1371, 772-1208, 786-1165, 786-1166, 789-1046, 791-1596, 808-1327, 810-1069, 817-983, 828-1087, 832-1356, 847-1073, 856-1242, 858-1502, 887-1414, 891-1713, 893-1509, 902-1384, 911-1259, 911-1261, 916-1378, 921-1691, 926-1398, 926-1559, 929-1179, 937-1459, 946-1226, 946-1232, 949-1123, 958-1138, 966-1483, 969-1434, 972-1670, 973-1135, 980-1579, 995-1266, 1002-1600, 1042-1457, 1047-1525, 1050-1700, 1064-1641, 1065-1261, 1065-1643, 1069-1660, 1079-1261, 1085-1338, 1135-1250, 1135-1261, 1142-1754, 1159-1491, 1183-1793, 1185-1998, 1196-1731, 1210-1810, 1213-1883, 1214-1759, 1218-1857, 1229-1938, 1234-1803, 1234-1870, 1234-1955, 1238-1976, 1252-1990, 1273-1982, 1279-1800, 1282-1852, 1284-1558, 1284-1996, 1296-2012, 1299-1959, 1303-1605, 1309-1849, 1330-1799, 1339-1799, 1339-1805, 1346-1769, 1348-1799, 1350-1590, 1350-1799, |
| 49 cont | 1362-1611, 1362-1738, 1363-1713, 1365-1851, 1366-2006, 1381-1628, 1390-1676, 1409-1653, 1411-1892, 1414-1800, 1431-1799, 1435-1574, 1440-2391, |
| 49 cont | 1457-1910, 1475-1997, 1488-1968, 1489-1844, 1591-1786, 1623-1879, 1670-1889, 1817-1979 |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 50/7503606CB1/ 4550 | 1-4550, 336-877, 339-697, 339-832, 372-598, 423-1222, 433-1043, 447-1166, 523-1189, 543-1190, 566-1157, 567-1287, 577-1174, 617-1261, 625-1258, 651-1136, 653-1069, 664-1262, 676-1444, 681-1444, 703-1261, 721-1261, 762-1275, 770-1472, 803-1288, 806-1277, 827-1265, 828-1305, 832-1275, 843-1275, 869-1476, 892-1291, 892-1573, 902-1417, 931-1590, 943-1274, 944-1300, 945-1275, 987-1684, 1005-1572, 1008-1444, 1041-1525, 1050-1503, 1057-1277, 1092-1563, 1147-1275, 1194-1275, 1317-1684, 3286-4001, 3292-3748, 3292-3753, 3362-3779, 3427-4120, 3514-4104 |
| 51/7500216CB1/ 2727 | 1-2727, 300-417, 311-648, 676-1247, 709-1514, 717-1373, 718-1172, 759-1320, 765-1402, 780-1017, 781-1301, 783-1288, 819-1012, 822-1515, 853-1433, 890-1533, 911-1116, 911-1201, 934-1533, 936-1227, 936-1390, 936-1429, 936-1460, 936-1521, 936-1544, 936-1551, 936-1574, 936-1599, 936-1608, 936-1620, 937-1587, 939-1462, 939-1550, 951-1512, 958-1605, 964-1625, 966-1670, 977-1142, 1008-1563, 1010-1653, 1057-1667, 1062-1345, 1063-1199, 1083-1322, 1083-1558, 1084-1494, 1087-1697, 1088-1648, 1098-1637, 1135-1310, 1135-1492, 1142-1308, 1160-1523, 1201-1387, 1202-1447, 1395-1667, 1415-1549, 1443-1699, 1553-1987, 1563-1752, 1598-1796, 1695-1738, 1713-1965, 1736-1893, 1788-1831, 1936-2407, 1950-2573, 1963-2524, 2077-2682, 2093-2411, 2093-2527, 2100-2531, 2106-2323, 2134-2528, 2136-2525, 2169-2525, 2170-2514, 2184-2525, 2185-2410, 2190-2526, 2198-2526, 2205-2525, 2209-2531, 2240-2525, 2241-2528, 2252-2471, 2285-2518, 2287-2523, 2292-2527, 2313-2581, 2317-2515, 2390-2525, 2395-2652, 2476-2727, 2525-2727 |
| 52/7099880CB1/ 4013 | 1-586, 7-739, 12-566, 12-628, 12-2328, 25-656, 40-731, 56-773, 57-535, 70-653, 70-794, 84-608, 84-710, 85-781, 651-1359, 680-1285, 683-1211, 683-1364, 688-1239, 855-1277, 1149-1457, 1181-1733, 1195-1416, 1195-1735, 1245-1733, 1254-1802, 1254-1889, 1365-2085, 1424-1689, 1481-1739, 1513-1816, 1513-2049, 2034-2484, 2034-2666, 2138-2614, 2144-2594, 2144-2775, 2183-2634, 2183-2813, 2239-2581, 2239-2706, 2239-2803, 2247-2842, 2377-2644, 2377-2696, 2406-2912, 2430-2957, 2439-3031, 2462-3049, 2498-2748, 2514-3209, 2563-3230, 2570-3061, 2576-3179, 2640-2891, 2650-3265, 2701-3399, 2759-2902, 2812-3487, 2914-3445, 2941-3663, 3010-3606, 3020-3549, 3020-3553, 3020-3558, 3020-3562, 3020-3570, 3020-3574, 3020-3576, 3020-3638, 3035-3576, 3044-3725, 3084-3707, 3143-3722, 3169-3760, 3186-3686, 3236-3460, 3253-3498, 3254-3708, 3287-3505, 3296-3873, 3298-3939, 3299-4013, 3300-3637, 3302-3762, 3336-3845, 3348-3646 |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 53/871513CB1/ 2768 | 1-648, 68-697, 89-326, 100-635, 114-384, 114-544, 117-738, 118-363, 118-490, 118-565, 118-580, 119-437, 120-398, 122-399, 124-404, 124-414, 125-220, 125-402, 125-605, 129-581, 136-377, 151-422, 152-467, 155-873, 166-626, 174-672, 180-711, 189-769, 192-422, 197-434, 197-477, 197-672, 197-699, 204-494, 218-687, 221-687, 227-687, 233-687, 235-687, 256-690, 267-392, 270-628, 285-692, 295-562, 311-481, 335-684, 345-690, 459-662, 466-1054, 551-687, 578-876, 581-854, 601-881, 656-1253, 681-957, 706-1004, 711-991, 715-1011, 716-1310, 718-970, 720-1264, 765-1084, 803-1028, 803-1067, 803-1557, 838-1087, 840-1083, 840-1108, 890-1140, 934-1202, 957-1585, 959-1174, 1085-1370, 1404-1680, 1421-1655, 1421-1680, 1476-1661, 1553-2171, 1555-2011, 1692-1969, 1723-2428, 1826-2404, 1874-2393, 1893-2127, 1933-2405, 1943-2405, 1945-2118, 2056-2347, 2060-2568, 2060-2768, 2093-2405, 2107-2320, 2121-2766 |
| 54/8057640CB1/ 5738 | 1-5501, 3145-3885, 3190-3878, 3190-3879, 3395-4014, 3406-4085, 3424-4019, 3461-4013, 3497-3981, 3532-4093, 3580-3956, 3609-4065, 3650-4213, 3671-3946, 3716-3964, 3717-4358, 3720-3975, 3752-4114, 3779-4470, 3789-4041, 3792-4015, 3792-4317, 3801-4374, 3803-4316, 3803-4358, 3810-4119, 3814-4513, 3817-4392, 3818-4304, 3820-4407, 3820-4487, 3822-4085, 3826-4109, 3832-4028, 3832-4295, 3846-4326, 3854-4032, 3854-4325, 3864-4102, 3879-4267, 3885-4434, 3895-4524, 3896-4326, 3905-4160, 3906-4348, 3914-4285, 3916-4176, 3928-4195, 3939-4617, 3943-4617, 3947-4519, 3953-4535, 3954-4617, 3969-4296, 3974-4208, 3974-4617, 3993-4627, 3994-4532, 3998-4181, 3998-4269, 4005-4617, 4016-4558, 4018-4617, 4021-4271, 4023-4617, 4025-4318, 4036-4303, 4038-4509, 4045-4515, 4048-4469, 4050-4169, 4053-4355, 4054-4617, 4055-4294, 4064-4544, 4069-4367, 4084-4320, 4092-4331, 4105-4624, 4107-4384, 4117-4359, 4135-4617, 4143-4617, 4145-4617, 4149-4627, 4153-4617, 4155-4411, 4166-4617, 4170-4618, 4175-4312, 4179-4548, 4190-4291, 4210-4614, 4221-4308, 4254-4608, 4259-4627, 4267-4542, 4267-4549, 4364-4617, 4516-4617, 4629-5181, 4665-4770, 4665-4778, 4699-5249, 4983-5738 |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 55/7505913CB1/ 1551 | 1-246, 1-314, 1-1551, 6-286, 7-294, 43-654, 59-265, 59-320, 70-815, 79-381, 80-320, 80-460, 80-516, 80-660, 80-708, 81-278, 81-292, 81-335, 81-350, 81-422, 81-639, 81-687, 81-846, 81-930, 82-289, 83-738, 84-714, 90-417, 91-381, 97-360, 99-412, 101-357, 104-335, 104-499, 108-732, 109-366, 111-347, 111-497, 113-261, 113-378, 113-867, 114-348, 114-350, 114-557, 114-636, 114-665, 114-754, 114-804, 114-968, 116-321, 116-325, 116-350, 116-383, 116-401, 116-581, 117-414, 117-905, 118-299, 118-331, 118-364, 118-394, 119-276, 119-360, 119-362, 119-375, 119-406, 119-684, 119-687, 120-342, 120-598, 122-909, 126-909, 129-390, 130-401, 136-579, 136-580, 139-720, 148-706, 167-464, 193-740, 197-476, 197-597, 210-787, 220-403, 234-646, 269-565, 280-489, 285-456, 299-1007, 303-579, 303-969, 313-423, 323-915, 330-915, 338-462, 344-924, 388-597, 388-607, 390-615, 390-1015, 391-663, 396-660, 406-1030, 411-695, 426-826, 426-920, 426-973, 426-1035, 430-1007, 431-702, 431-979, 431-1025, 438-729, 439-600, 458-843, 476-722, 508-763, 533-626, 534-997, 535-997, 535-1058, 540-780, 544-810, 546-1112, 557-1026, 578-840, 581-876, 593-862, 597-813, 599-771, 599-773, 600-1024, |
| 55 cont | 604-816, 620-911, 624-953, 624-1229, 625-844, 631-901, 632-912, 632-947, 634-911, 637-1144, 640-930, 641-821, 650-934, 659-961, 661-932, 665-1188, 667-881, 677-1183, 677-1242, 678-1029, 683-1002, 691-881, 705-957, 705-967, 716-982, 727-827, 728-1182, 783-900, 798-1175, 799-1175, 803-1183, 818-1225, 828-1121, 828-1122, 875-1129, 883-1205, 904-1184, 918-1206, 944-1214, 969-1172, 1018-1174, 1038-1153, 1160-1534, 1260-1540, 1261-1551, 1289-1482, 1341-1465, 1365-1534, 1367-1551 |
| 56/7510292CB1/ 4254 | 1-118, 1-482, 8-478, 9-472, 9-474, 9-4254, 176-467, 176-481, 176-745, 184-859, 308-859, 361-900, 454-1024, 563-1058, 604-851, 618-1034, 672-1122, 784-887, 813-1086, 876-1543, 886-1543, 958-1543, 960-1543, 984-1543, 993-1642, 1008-1295, 1015-1543, 1017-1535, 1017-1543, 1019-1541, 1034-1543, 1055-1869, 1070-1336, 1086-1543, 1100-1767, 1119-1530, 1119-1534, 1119-1543, 1126-1250, 1181-1543, 1192-1281, 1197-1543, 1281-1768, 1322-1543, 1340-1543, 1343-2063, 1359-1543, 1386-2007, 1388-1543, 1395-1769, 1407-1543, 1409-1543, 1426-2032, 1475-1543, 1488-1869, 1785-1968, 1785-2071, 1791-2094, 1833-2071, 1834-2057, 1834-2116, 1834-2120, 1834-2130, 1834-2133, 1834-2261, 1834-2289, 1834-2379, 1834-2404, 1834-2504, 1834-2515, 1857-2129, 1920-2376, 1922-2617, 2003-2511, 2092-2322, 2102-2390, 2102-2548, 2107-2409, 2134-2393, 2173-2430, 2287-2409, 2453-2736, 2453-2922, 2622-2827, 2642-2819, 2681-3488, 2819-3389, 2870-3417, 2874-3395, 2882-3444, 2886-3779, 2896-3401, 2896-3415, 2896-3421, 2896-3444, 2927-3763, 2941-3219, 2975-3231, 2978-3425, 2984-3220, 2998-3268, 3014-3485, 3014-3546, 3021-3876, 3026-3323, 3030-3877, 3057-3608, 3060-3336, 3062-3382, 3062-3383, 3067-3589, 3075-3553, 3085-3569, 3099-3869, 3104-3380, 3105-3411, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 56 cont | 3131-3398, 3131-3773, 3132-4058, 3133-3797, 3137-3693, 3160-3400, 3161-3723, 3164-3620, 3177-3422, 3204-3572, 3204-3833, 3211-3438, 3213-3496, 3227-3521, 3247-3526, 3250-3572, 3256-4026, 3266-3518, 3276-3567, 3287-3564, 3305-3912, 3326-3552, 3343-3643, 3348-3605, 3360-3510, 3361-3904, 3379-3686, 3384-3623, 3406-3659, 3416-3710, 3430-3828, 3437-3636, 3441-3692, 3443-3645, 3462-3736, 3475-4043, 3481-3771, 3485-3795, 3485-3881, 3486-3894, 3494-4041, 3499-3756, 3499-3803, 3511-3750, 3514-3795, 3519-3788, 3539-3791, 3550-3877, 3551-3734, 3555-3791, 3555-3828, 3565-3831, 3565-3905, 3573-4128, 3575-3822, 3577-4186, 3585-4231, 3593-3736, 3594-3816, 3607-3852, 3608-4150, 3614-3787, 3615-3934, 3625-4031, 3631-4235, 3636-3897, 3642-4130, 3649-3905, 3652-4115, 3657-4143, 3667-3892, 3669-3950, 3672-3901, 3672-3906, 3675-4213, 3676-3898, 3676-4127, 3686-4192, 3687-3918, 3724-4253, 3734-4132, 3736-3989, 3736-4023, 3748-4224, 3749-3964, 3751-4254, 3754-4043, 3763-3989, 3763-4238, 3763-4245, 3767-4233, 3771-4237, 3778-4220, 3785-4009, 3785-4041, 3788-4236, 3796-4019, 3797-3946, 3799-4202, 3805-4232, 3811-4197, 3827-4235, 3829-4231, 3835-4012, 3835-4113, |
| 56 cont | 3848-4238, 3848-4254, 3852-4254, 3855-4096, 3855-4131, 3881-4238, 3890-4238, 3894-4238, 3897-4191, 3903-4228, 3903-4243, 3926-4233, 3928-4254, 3936-4196, 3944-4166, 3951-4254, 3955-4248, 3956-4238, 3956-4241, 3966-4238, 3974-4254, 3978-4245, 3991-4240, 3991-4254, 3992-4234, 4001-4234, 4005-4254, 4015-4253, 4016-4254, 4025-4244, 4040-4238, 4062-4233, 4067-4252, 4069-4247, 4069-4254, 4083-4178, 4083-4254, 4110-4245, 4120-4254 |
| 57/7504669CB1/ 1509 | 1-554, 9-1509, 680-992, 760-1017, 775-982, 775-1253, 776-1032, 782-1398, 789-1060, 792-1234, 799-1044, 819-1051, 820-1398, 824-1060, 856-961, 862-1162, 868-1143, 873-1079, 873-1136, 877-1130, 880-1153, 881-1202, 885-1075, 885-1133, 888-1132, 900-1109, 904-1359, 916-1172, 933-1035, 935-1383, 939-1108, 942-1386, 995-1245, 1029-1205, 1076-1398, 1257-1379, 1257-1395, 1257-1398, 1257-1403, 1257-1404, 1258-1398, 1271-1505 |
| 58/7509266CB1/ 2439 | 1-256, 1-327, 1-335, 1-1130, 1-2439, 23-303, 25-269, 133-973, 170-934, 170-1015, 411-508, 425-963, 499-948, 593-1192, 704-1395, 746-1229, 752-1515, 762-1317, 808-1505, 850-1548, 882-1612, 919-1433, 1054-1319, 1289-1912, 1321-1403, 1377-1760, 1407-1988, 1410-2103, 1453-2398, 1458-2401, 1465-2396, 1480-2401, 1499-2401, 1508-2051, 1519-2402, 1529-2401, 1547-2439, 1554-2110, 1559-2439, 1565-2083, 1574-2401, 1586-2168, 1598-2439, 1599-2439, 1615-2117, 1642-2019, 1645-2398, 1645-2439, 1693-1933, 1706-2439, 1724-2355, 1735-2401, 1744-2402, 1851-2155, 2088-2366 |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 59/7509288CB1/ 2440 | 1-252, 1-256, 1-335, 1-2440, 23-303, 25-269, 133-973, 170-934, 170-943, 366-1040, 366-1051, 367-962, 411-508, 425-963, 430-969, 499-948, 551-1115, 583-1254, 593-1192, 603-1257, 614-1237, 631-1272, 679-989, 694-1189, 704-1395, 721-1257, 746-1229, 752-1395, 762-1317, 808-1449, 850-1549, 882-1613, 988-1626, 1018-1626, 1033-1626, 1054-1319, 1055-1626, 1058-1626, 1068-1626, 1109-1601, 1110-1408, 1115-1266, 1212-1716, 1221-1433, 1289-1552, 1289-1730, 1289-1805, 1289-1856, 1289-1859, 1289-1870, 1289-1913, 1289-1920, 1321-1403, 1363-1869, 1370-1873, 1377-1617, 1377-1761, 1407-1989, 1410-2104, 1414-2027, 1453-2399, 1465-2397, 1468-2033, 1481-2402, 1498-2061, 1500-2402, 1509-2052, 1548-2402, 1548-2440, 1555-1918, 1555-2111, 1560-2402, 1560-2440, 1562-2402, 1566-2084, 1566-2158, 1575-2440, 1579-2440, 1587-2169, 1595-2144, 1596-2437, 1599-2440, 1600-2440, 1604-2440, 1616-2118, 1621-2403, 1623-2402, 1643-2020, 1644-2397, 1645-2397, 1645-2402, 1646-2399, 1646-2440, 1694-1934, 1702-2402, 1705-2402, 1707-2440, 1716-2402, 1725-2356, 1736-2402, 1741-2256, 1745-2403, 1750-2324, 1852-2156, 1870-2402, 1888-2140, 1889-2397, 1905-2403, 2089-2367, 2095-2378, 2095-2385, 2095-2390, 2097-2411 |
| 60/7510212CB1/ 2580 | 1-428, 1-2580, 49-656, 56-617, 57-596, 57-736, 58-524, 58-663, 61-590, 61-632, 61-704, 62-680, 63-570, 66-654, 70-927, 72-526, 75-775, 82-683, 103-745, 138-753, 160-608, 168-387, 169-461, 204-390, 210-782, 213-436, 213-451, 255-621, 268-519, 269-857, 275-386, 280-538, 282-571, 290-972, 297-757, 300-545, 314-611, 317-523, 317-572, 317-576, 326-561, 328-584, 329-507, 329-646, 336-881, 348-643, 348-753, 358-515, 360-1020, 362-468, 366-733, 384-831, 396-619, 398-1055, 399-672, 399-849, 399-860, 399-882, 399-955, 399-983, 399-988, 399-1015, 399-1070, 399-1084, 399-1087, 399-1114, 399-1129, 399-1133, 399-1186, 399-1204, 399-1236, 404-1098, 410-727, 432-1021, 445-661, 445-719, 446-1140, 446-1207, 447-694, 449-588, 449-719, 454-700, 472-1140, 475-595, 483-732, 483-1044, 491-725, 497-584, 501-790, 501-802, 503-765, 504-743, 513-809, 523-718, 525-1103, 528-1051, 532-770, 535-756, 535-793, 551-1473, 553-846, 566-824, 568-788, 568-843, 572-952, 572-976, 574-1388, 578-1262, 579-847, 585-1105, 589-1156, 591-893, 593-843, 595-849, 596-1397, 600-1182, 606-1161, 612-890, 616-862, 616-877, 616-894, 617-1080, 617-1228, 619-1306, 630-828, 630-1249, 634-918, 640-972, |
| 60 cont | 640-1020, 640-1057, 640-1094, 640-1098, 640-1176, 640-1225, 640-1235, 641-870, 643-879, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 60 cont | 643-883, 646-1039, 652-1310, 655-912, 656-880, 658-897, 658-1473, 659-1000, 660-941, 664-891, 666-799, 666-1362, 667-1106, 671-972, 675-972, 676-857, 678-1211, 678-1292, 681-1113, 681-1114, 684-1237, 685-1388, 688-774, 692-968, 692-1340, 696-1284, 698-958, 699-998, 699-1341, 708-987, 712-1265, 712-1313, 721-1416, 721-1468, 721-1470, 724-1276, 727-926, 730-981, 736-1220, 744-1180, 748-1041, 752-1152, 753-1033, 753-1043, 759-1044, 762-1022, 766-1398, 771-1373, 773-1045, 775-922, 779-1470, 780-1401, 781-1348, 782-1100, 782-1157, 782-1266, 782-1290, 782-1309, 782-1321, 785-1353, 787-1044, 789-1369, 792-985, 792-1021, 795-1139, 800-894, 800-1012, 800-1044, 800-1101, 801-1329, 801-1330, 803-1071, 803-1462, 807-1275, 809-960, 812-1053, 814-1073, 815-1473, 816-1375, 816-1473, 818-1053, 819-1179, 819-1431, 824-1091, 824-1097, 826-1112, 828-1082, 835-1105, 839-1116, 839-1470, 844-1100, 847-1121, 853-1113, 855-1131, 855-1132, 860-1136, 867-1080, 869-1355, 872-1076, 876-1113, 880-1099, 880-1142, 883-1462, 885-1155, 888-1433, 889-1462, 892-1352, 892-1473, 900-1352, 906-1412, 908-1433, 910-1364, 911-1212, 911-1409, 911-1410, 912-1336, 921-1425, |
| 60 cont | 923-1458, 927-1171, 928-1174, 931-1030, 931-1169, 931-1222, 932-1444, 933-1225, 933-1318, 937-1473, 941-1130, 942-1213, |
| 60 cont | 942-1390, 943-1191, 943-1206, 948-1155, 948-1191, 948-1473, 952-1410, 954-1256, 956-1171, 956-1249, 960-1221, 969-1236, 976-1246, 976-1252, 978-1350, 980-1178, 985-1206, 986-1190, 986-1204, 986-1275, 987-1238, 987-1263, 989-1285, 989-1330, 993-1470, 996-1254, 996-1272, 999-1117, 1006-1244, 1006-1434, 1006-1473, 1007-1246, 1007-1254, 1007-1465, 1009-1258, 1009-1298, 1010-1113, 1011-1473, 1016-1473, 1020-1473, 1027-1473, 1033-1343, 1035-1450, 1038-1296, 1038-1473, 1039-1290, 1039-1473, 1040-1308, 1048-1332, 1048-1343, 1048-1359, 1049-1302, 1049-1343, 1049-1473, 1056-1471, 1058-1265, 1060-1469, 1061-1269, 1061-1332, 1061-1336, 1063-1318, 1063-1342, 1063-1353, 1070-1311, 1070-1470, 1071-1344, 1071-1473, 1072-1334, 1072-1473, 1075-1473, 1076-1338, 1079-1473, 1085-1473, 1086-1312, 1086-1340, 1086-1368, 1087-1473, 1089-1343, 1089-1368, 1090-1336, 1090-1351, 1090-1377, 1092-1385, 1093-1260, 1094-1338, 1094-1360, 1094-1473, 1096-1248, 1096-1351, 1098-1369, 1103-1354, 1103-1473, 1112-1333, 1115-1350, 1124-1143, 1126-1241, 1132-1377, 1134-1356, 1134-1358, 1134-1366, 1134-1367, 1136-1391, 1141-1417, 1148-1294, 1148-1414, 1149-1455, 1152-1466, 1154-1463, 1155-1339, 1156-1259, 1156-1404, |
| 60 cont | 1164-1473, 1168-1435, 1171-1473, 1172-1441, 1185-1473, 1186-1473, 1195-1332, 1195-1407, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 60 cont | 1205-1468, 1206-1471, 1208-1473, 1220-1473, 1239-1804, 1258-1463, 1272-1473, 1281-1473, 1342-1473, 1379-1473, 1381-1473, 1386-1473, 1398-1473, 1400-1473, 1404-1473, 1406-1473, 1422-1473, 1433-1473, 1435-1473, 1441-1473, 1442-1473, 1446-1473, 1476-1497, 1476-1500, 1476-1502, 1476-1504, 1476-1507, 1476-1521, 1476-1523, 1476-1525, 1476-1531, 1476-1534, 1476-1552, 1476-1553, 1476-1555, 1476-1556, 1476-1558, 1476-1559, 1476-1560, 1476-1561, 1476-1563, 1477-1563, 1478-1506, 1478-1545, 1478-1548, 1478-1552, 1478-1558, 1478-1560, 1478-1563, 1488-1563, 1490-1719, 1494-1563, 1494-1764, 1496-1563, 1512-1563, 1523-1563, 1525-1563, 1531-1563, 1532-1563, 1536-1791, 1565-1774 |
| 61/7510504CB1/ 8029 | 1-678, 206-478, 206-8029, 259-548, 261-536, 278-349, 286-512, 286-664, 286-691, 297-814, 303-691, 466-1161, 584-1139, 646-1062, 658-1294, 662-1130, 711-1277, 719-1372, 820-1367, 820-1383, 820-1406, 843-1418, 855-1431, 864-1431, 869-1281, 873-1487, 930-1222, 930-1400, 930-1516, 930-1519, 930-1592, 930-1622, 930-1681, 930-1748, 930-1769, 1006-1461, 1036-1629, 1070-1337, 1313-1642, 1314-1591, 1631-1911, 1890-2289, 2123-2390, 2213-2376, 2214-2554, 2260-2529, 2264-2925, 2387-2925, 2406-2673, 2453-2934, 2455-2913, 2464-2900, 2518-2934, 2562-2904, 2700-2916, 2763-2887, 2889-3433, 2927-3431, 2936-3470, 3118-3830, 3244-3830, 3264-3830, 3548-4142, 3944-4225, 4166-4459, 4320-4914, 4376-4585, 4586-5188, 4716-5375, 4839-5135, 5021-5609, 5102-5733, 5188-5446, 5188-5473, 5188-5795, 5287-5894, 5435-6042, 5745-6376, 5802-6434, 5866-6472, 5978-6504, 6109-6450, 6150-6424, 6150-6698, 6201-6833, 6227-6388, 6576-7148 |
| 62/7510587CB1/ 1868 | 1-281, 1-522, 1-1696, 43-755, 43-937, 43-938, 43-970, 43-993, 63-269, 63-694, 513-603, 533-595, 584-1345, 891-1354, 923-1704, 927-1712, 1024-1318, 1072-1684, 1098-1403, 1099-1546, 1316-1385, 1467-1697, 1594-1868 |
| 63/7510684CB1/ 1531 | 1-1523, 9-605, 223-515, 223-550, 223-568, 223-933, 223-960, 224-567, 227-781, 227-1020, 273-665, 274-1139, 340-1008, 400-494, 446-1375, 478-995, 537-1143, 559-1375, 607-1375, 675-1214, 799-1197, 912-1476, 918-1177, 1000-1514, 1059-1502, 1075-1526, 1092-1522, 1093-1314, 1100-1517, 1111-1520, 1117-1494, 1127-1517, 1144-1521, 1150-1511, 1150-1517, 1387-1526, 1411-1486, 1411-1526, 1411-1531 |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 64/7510697CB1/ 4220 | 1-487, 1-3559, 9-487, 13-482, 14-481, 14-482, 19-499, 181-816, 189-864, 313-864, 317-668, 366-905, 459-1029, 491-1063, 609-856, 673-1047, 850-1548, 854-1130, 876-1548, 939-1463, 963-1548, 998-1647, 1013-1300, 1075-1341, 1286-1773, 1371-1881, 1391-2012, 1400-1849, 1431-2037, 1674-2119, 1796-2119, 1839-2062, 1839-2266, 1990-2514, 2008-2516, 2097-2327, 2107-2395, 2111-2635, 2112-2414, 2139-2398, 2178-2435, 2375-2672, 2397-2936, 2398-2983, 2424-3008, 2465-2733, 2507-3042, 2507-3057, 2516-2786, 2547-2803, 2602-3200, 2675-3103, 2720-3335, 2736-2924, 2782-3020, 2782-3352, 2795-3388, 2804-3066, 2821-3146, 2832-3063, 2832-3167, 2833-3380, 2837-3358, 2845-3407, 2859-3407, 2890-3141, 2904-3182, 2938-3194, 2938-3306, 2947-3183, 2961-3231, 2975-3244, 2975-3245, 2977-3434, 2977-3448, 2989-3286, 3023-3299, 3030-3552, 3038-3516, 3048-3532, 3068-3374, 3096-3627, 3100-3656, 3124-3686, 3167-3594, 3174-3401, 3176-3594, 3210-3489, 3213-3594, 3229-3481, 3239-3526, 3250-3527, 3268-3875, 3306-3606, 3311-3568, 3320-3558, 3324-3867, 3342-3649, 3345-3888, 3347-3586, 3367-3622, 3369-3622, 3379-3673, 3404-3655, 3406-3608, 3438-4006, 3448-3758, 3449-3841, 3457-4004, 3462-3719, 3462-3766, 3474-3713, 3482-3751, |
| 64 cont | 3502-3754, 3513-3840, 3514-3697, 3518-3754, 3518-3838, 3522-3784, 3528-3789, 3528-3794, 3536-4091, 3538-3785, 3557-3779, 3570-3815, 3571-4113, 3578-4019, 3588-4096, 3599-3860, 3605-4189, 3612-3868, 3615-4199, 3630-3855, 3632-3913, 3635-3864, 3635-3869, 3635-4194, 3638-4176, 3639-3861, 3639-4199, 3649-4155, 3650-3881, 3687-4216, 3688-4220, 3697-4001, 3697-4215, 3699-3952, 3699-3986, 3711-4187, 3712-3927, 3714-4220, 3717-4006, 3726-3952, 3726-4201, 3726-4208, 3730-4196, 3734-4200, 3741-4183, 3751-4199, 3759-4197, 3759-4220, 3762-4165, 3768-4191, 3768-4195, 3768-4220, 3774-4160, 3786-4220, 3790-4198, 3792-4194, 3798-3975, 3798-4076, 3811-4201, 3811-4220, 3815-4217, 3816-4206, 3844-4201, 3853-4201, 3857-4201, 3860-4154, 3863-4153, 3889-4196, 3899-4201, 3907-4129, 3914-4219, 3918-4211, 3919-4201, 3919-4217, 3929-4201, 3941-4208, 3954-4203, 3954-4220, 3964-4197, 3968-4220, 3978-4216, 3979-4220, 4046-4220 |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 65/7761337CB1/ 5353 | 1-4789, 626-677, 914-977, 1077-1629, 1158-1184, 1173-1391, 1174-1377, 1263-1815, 1267-1629, 1280-1331, 1286-1349, 1287-1323, 1300-1901, 1323-1817, 1336-1391, 1429-1484, 1470-1888, 1545-1603, 1545-1629, 1546-1629, 1546-1749, 1551-1603, 1555-1629, 1558-2133, 1560-1802, 1596-2079, 1596-2273, 1604-1849, 1604-1858, 1615-1670, 1622-1866, 1623-2285, 1628-2042, 1648-1816, 1650-1909, 1708-1763, 1727-2312, 1731-1815, 1742-2441, 1945-2589, 1962-2377, 1989-2614, 2005-2637, 2035-2431, 2049-2606, 2050-2425, 2087-2643, 2116-2360, 2126-2602, 2159-2382, 2173-2440, 2287-2824, 2299-2878, 2309-2972, 2393-2988, 2412-2843, 2467-2770, 2478-2883, 2539-2834, 2583-3065, 2595-2903, 2595-3120, 2595-3422, 2617-3166, 2656-3230, 2806-3235, 2836-3497, 2858-3191, 2859-3536, 2888-3483, 2905-3560, 2923-3208, 2926-3214, 2931-3225, 2940-3164, 2985-3210, 3016-3268, 3227-3712, 3236-3711, 3238-3677, 3422-3563, 3460-3634, 3471-4049, 3520-4014, 3523-4090, 3560-4144, 3651-4176, 3652-4230, 3657-4262, 3666-4126, 3666-4239, 3671-4115, 3688-4214, 3707-4294, 3714-4224, 3725-4329, 3742-4345, 3766-4265, 3772-4448, 3788-4373, 3792-4439, 3795-4408, 3797-4481, 3825-4402, 3836-4382, 3836-4429, 3839-4550, 3862-4185, 3862-4364, 3865-4121, |
| 65 cont | 3886-4434, 3886-4469, 3888-4152, 3888-4172, 3893-4183, 3895-4345, 3904-4408, 3918-4595, 3925-4625, 3930-4020, 3941-4402, 3949-4450, 3963-4266, 3968-4215, 3971-4250, 3971-4665, 3974-4470, 3974-4536, 3976-4222, 3978-4236, 3982-4710, 3984-4222, 3994-4288, 4004-4721, 4006-4253, 4015-4504, 4018-4618, 4030-4622, 4043-4253, 4043-4629, 4053-4698, 4074-4537, 4076-4674, 4078-4314, 4078-4740, 4080-4519, 4080-4613, 4082-4606, 4082-4650, 4087-4744, 4088-4648, 4096-4341, 4096-4346, 4104-4744, 4107-4728, 4108-4331, 4121-4784, 4122-4638, 4122-4761, 4130-4652, 4138-4637, 4139-4691, 4155-4374, 4157-4784, 4163-4707, 4178-4405, 4178-4694, 4180-4689, 4183-4851, 4196-4864, 4216-4740, 4231-4684, 4242-4914, 4245-4689, 4245-4740, 4249-4926, 4253-4560, 4255-4523, 4260-4513, 4263-4781, 4265-4969, 4269-4893, 4271-4534, 4272-4918, 4295-4782, 4298-4598, 4311-4957, 4313-4567, 4314-4586, 4314-4617, 4318-5004, 4320-4990, 4325-4938, 4330-5022, 4331-5002, 4332-4592, 4332-4805, 4334-4588, 4336-4635, 4338-4587, 4348-4605, 4349-4814, 4355-4558, 4357-4643, 4358-4881, 4369-4830, 4372-4929, 4375-4830, 4376-5017, 4378-4970, 4385-4848, 4386-4662, 4387-5076, 4389-4642, 4393-4644, 4400-5112, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 65 cont | 4404-4685, 4407-4916, 4409-4875, 4411-4848, 4422-4634, 4422-4643, 4422-4659, 4423-5111, 4426-4957, 4430-4660, 4434-4766, 4439-4657, 4439-4884, 4439-4890, 4439-4907, 4439-4916, 4439-4929, 4439-4960, 4439-4961, 4439-4969, 4439-5062, 4442-5056, 4446-4910, 4446-4983, 4449-4702, 4453-5084, 4461-5013, 4464-4961, 4465-4734, 4466-4720, 4467-5147, 4482-5103, 4482-5178, 4496-4792, 4498-5042, 4500-5140, 4504-5091, 4510-4982, 4512-5213, 4517-5063, 4519-5117, 4524-4781, 4530-4796, 4535-5213, 4535-5219, 4536-4777, 4538-5098, 4538-5142, 4538-5205, 4542-5129, 4544-4822, 4548-5186, 4560-5118, 4564-5062, 4564-5132, 4565-5240, 4576-5224, 4586-5336, 4587-5218, 4594-5277, 4612-5224, 4633-4766, 4633-4772, 4641-5222, 4648-5225, 4649-5325, 4658-4841, 4666-4927, 4669-5335, 4675-4925, 4683-4967, 4689-5333, 4707-5233, 4714-5207, 4719-5327, 4721-5340, 4721-5352, 4723-5300, 4724-5283, 4724-5340, 4724-5353, 4730-5219, 4731-5309, 4734-4967, 4738-5300, 4740-5284, 4745-5326, 4755-5286, 4758-4931, 4758-5235, 4765-5283, 4768-5335, 4768-5338, 4773-5227, 4777-5303, 4787-5306, 4792-4979, 4801-5022, 4806-5310, 4822-5301, 4831-5340, 4842-5076, 4843-5082, 4862-5121, 4890-5135, |
| 65 cont | 4945-5164, 4967-5180 |
| 66/7503666CB1/ 3126 | 1-2407, 1-3120, 260-764, 274-850, 275-522, 275-778, 275-969, 275-1023, 275-1196, 278-438, 279-704, 279-838, 280-1201, 281-548, 281-554, 285-762, 285-811, 285-871, 286-933, 287-621, 287-797, 287-871, 289-855, 289-962, 292-828, 293-927, 293-968, 300-417, 302-721, 302-857, 303-927, 304-590, 304-939, 304-945, 304-952, 305-906, 307-786, 311-648, 313-877, 315-878, 315-948, 319-849, 319-927, 324-951, 324-969, 329-632, 400-1135, 402-900, 410-1139, 454-1111, 466-1139, 469-1083, 472-1175, 484-931, 488-1069, 518-1133, 542-1095, 543-960, 543-1090, 550-1047, 566-1179, 584-1130, 602-1177, 629-1141, 633-1201, 634-1110, 718-1172, 780-1017, 781-1201, 811-1522, 819-1012, 911-1116, 911-1201, 977-1142, 1063-1199, 1084-1333, 1370-1637, 1374-1545, 1388-1905, 1415-2065, 1422-2080, 1429-2065, 1476-1725, 1664-2084, 1704-1928, 1743-2214, 1743-2215, 1786-2323, 1846-2077, 1847-2106, 1848-1938, 1848-1972, 1848-1985, 1848-1986, 1848-2000, 1848-2040, 1848-2083, 1848-2121, 1848-2122, 1848-2131, 1849-2050, 1850-1972, 1850-2131, 1864-2131, 1866-1898, 1866-1907, 1866-1917, 1866-1923, 1866-1935, 1866-1936, 1866-1945, 1866-1950, 1866-1957, 1866-1984, 1866-1990, |
| 66 cont | 1866-2013, 1866-2016, 1866-2028, 1866-2029, 1866-2042, 1866-2131, 1869-2131, 1875-2131, 1879-2093, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 66 cont | 1879-2389, 1879-2391, 1883-2430, 1884-2136, 1886-2042, 1891-2106, 1905-2488, 1906-2042, 1909-2031, 1914-2042, 1917-2143, 1917-2176, 1920-2131, 1922-2042, 1923-2482, 1924-2131, 1928-2457, 1935-2199, 1941-2158, 1941-2173, 1941-2224, 1943-2021, 1943-2042, 1943-2158, 1943-2177, 1943-2224, 1943-2528, 1972-2186, 1972-2224, 1977-2224, 1984-2202, 1991-2538, 1998-2224, 1999-2131, 2002-2124, 2007-2131, 2012-2224, 2015-2131, 2016-2224, 2017-2428, 2021-2222, 2026-2224, 2028-2224, 2036-2131, 2036-2224, 2052-2114, 2052-2224, 2052-2228, 2065-2228, 2070-2228, 2077-2224, 2084-2224, 2092-2470, 2095-2217, 2100-2305, 2106-2358, 2108-2228, 2109-2228, 2110-2224, 2111-2224, 2114-2228, 2119-2228, 2121-2228, 2129-2228, 2129-2286, 2145-2207, 2145-2224, 2158-2224, 2165-2224, 2170-2228, 2185-2224, 2193-2224, 2198-2224, 2202-2224, 2203-2228, 2343-2966, 2356-2917, 2470-3075, 2486-2804, 2486-2920, 2493-2924, 2499-2716, 2527-2921, 2529-2918, 2562-2918, 2563-2907, 2577-2918, 2578-2803, 2583-2919, 2591-2919, 2598-2918, 2602-2924, 2633-2918, 2634-2921, 2645-2864, 2678-2911, 2680-2916, 2685-2920, 2706-2974, 2710-2908, 2783-2918, 2788-3045, 2869-3122, 2918-3126 |
| 67/7503668CB1/ 3066 | 1-2347, 1-3060, 275-522, 275-702, 278-438, 281-548, 281-554, 287-621, 300-417, 304-590, 311-634, 329-632, 633-1079, 633-1154, 633-1421, 641-1280, 666-1227, 687-924, 688-1208, 690-1195, 726-919, 760-1340, 818-1023, 818-1108, 843-1134, 843-1297, 843-1336, 843-1367, 843-1428, 843-1451, 843-1458, 843-1481, 843-1506, 843-1515, 843-1527, 844-1494, 846-1369, 846-1457, 858-1419, 867-1137, 873-1577, 884-1049, 964-1574, 969-1252, 991-1401, 1005-1544, 1018-1675, 1020-1641, 1026-1137, 1027-1137, 1049-1215, 1055-1868, 1067-1430, 1077-1780, 1088-1838, 1109-1354, 1328-1845, 1355-2005, 1362-2020, 1369-2005, 1416-1665, 1604-2024, 1644-1868, 1683-2154, 1683-2155, 1726-2263, 1786-2017, 1787-2046, 1788-1878, 1788-1912, 1788-1925, 1788-1926, 1788-1940, 1788-1980, 1788-2023, 1788-2061, 1788-2062, 1788-2071, 1789-1990, 1790-1912, 1790-2071, 1804-2071, 1806-1838, 1806-1847, 1806-1857, 1806-1863, 1806-1875, 1806-1876, 1806-1885, 1806-1890, 1806-1897, 1806-1924, 1806-1930, 1806-1953, 1806-1956, 1806-1968, 1806-1969, 1806-1982, 1806-2071, 1809-2071, 1815-2071, 1819-2033, 1819-2329, 1819-2331, 1823-2370, 1824-2076, 1826-1982, 1831-2046, 1845-2428, 1846-1982, 1849-1971, 1854-1982, 1857-2083, 1857-2116, 1860-2071, 1862-1982, 1863-2422, 1864-2071, 1868-2397, 1875-2139, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 67 cont | 1881-1961, 1881-2098, 1881-2103, 1881-2113, 1881-2164, 1883-1961, 1883-1982, 1883-2098, 1883-2117, 1883-2164, 1883-2468, 1912-2126, 1912-2164, 1917-2164, 1924-2142, 1931-2478, 1938-2164, 1939-2071, 1942-2064, 1947-2071, 1952-2164, 1955-2071, 1956-2164, 1957-2224, 1957-2368, 1961-2162, 1966-2164, 1968-2164, 1976-2071, 1976-2164, 1992-2054, 1992-2164, 1992-2168, 2005-2168, 2010-2168, 2017-2164, 2024-2164, 2032-2410, 2035-2157, 2040-2245, 2046-2298, 2048-2168, 2049-2168, 2050-2164, 2051-2164, 2054-2168, 2059-2168, 2061-2168, 2069-2168, 2069-2226, 2085-2147, 2085-2164, 2098-2164, 2105-2164, 2110-2168, 2125-2164, 2133-2164, 2138-2164, 2142-2164, 2143-2168, 2283-2906, 2296-2857, 2410-3015, 2426-2744, 2426-2860, 2433-2864, 2439-2656, 2467-2861, 2469-2858, 2502-2858, 2503-2847, 2517-2858, 2518-2743, 2523-2859, 2531-2859, 2538-2858, 2542-2864, 2573-2858, 2574-2861, 2585-2804, 2618-2851, 2620-2856, 2625-2860, 2646-2914, 2650-2848, 2723-2858, 2728-2985, 2809-3062, 2858-3066 |
| 68/7503672CB1/ 3045 | 1-2326, 1-3039, 260-764, 274-850, 275-522, 275-778, 275-969, 275-1023, 275-1203, 278-438, 279-704, 279-838, 280-1219, 281-548, 281-554, 285-762, 285-811, 285-871, 286-933, 287-797, 287-871, 289-855, 289-962, 292-828, 293-927, 293-968, 300-417, 302-721, 302-857, 303-927, 304-590, 304-939, 304-945, 304-952, 305-906, 311-648, 313-877, 315-878, 319-849, 319-927, 324-951, 324-969, 329-632, 400-1135, 402-900, 410-1139, 454-1111, 466-1139, 469-1083, 472-1175, 484-931, 518-1133, 542-1095, 543-960, 543-1090, 550-1047, 566-1179, 590-1270, 602-1177, 629-1141, 633-1305, 634-1110, 676-1247, 709-1514, 712-1373, 718-1172, 759-1320, 780-1017, 781-1301, 783-1288, 853-1433, 911-1116, 911-1201, 936-1227, 936-1390, 936-1429, 936-1460, 936-1521, 936-1544, 936-1551, 936-1574, 936-1599, 936-1608, 936-1620, 937-1587, 939-1462, 939-1550, 951-1512, 966-1670, 977-1142, 1057-1667, 1062-1345, 1084-1494, 1098-1637, 1111-1768, 1113-1734, 1142-1308, 1148-1961, 1160-1523, 1170-1873, 1181-1931, 1201-1387, 1202-1447, 1421-1938, 1448-2098, 1455-2113, 1462-2098, 1509-1758, 1606-2290, 1653-1709, 1682-2224, 1697-2117, 1705-2183, 1708-2287, 1722-2232, 1737-1961, |
| 68 cont | 1836-2321, 1879-2110, 1881-1971, 1881-2005, 1881-2018, 1881-2019, 1881-2033, 1881-2073, 1883-2005, 1899-1931, 1899-1940, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 68 cont | 1899-1983, 1899-1984, 1899-2017, 1912-2126, 1917-2169, 1924-2073, 1942-2064, 1955-2073, 1962-2073, 1974-2054, 1974-2165, 1974-2166, 1976-2054, 1976-2165, 1976-2166, 2005-2166, 2010-2166, 2016-2449, 2035-2157, 2041-2691, 2055-2730, 2085-2147, 2085-2165, 2085-2167, 2085-2170, 2098-2170, 2103-2170, 2109-2166, 2134-2166, 2171-2837, 2228-2397, 2248-2719, 2262-2885, 2274-2447, 2275-2836, 2389-2994, 2405-2723, 2405-2839, 2412-2843, 2418-2635, 2446-2840, 2448-2837, 2481-2837, 2482-2826, 2496-2837, 2497-2722, 2502-2838, 2510-2838, 2517-2837, 2521-2843, 2552-2837, 2553-2840, 2564-2783, 2597-2830, 2599-2835, 2604-2839, 2625-2893, 2629-2827, 2702-2837, 2707-2964, 2788-3041, 2837-3045 |
| 69/6039650CB1/ 4170 | 1-4170, 1218-2248, 1236-2263, 1271-2228, 1296-1907, 1347-1913, 1432-2188, 1435-1955, 1478-1988, 1509-1949, 1527-1946, 1545-1959, 1551-2551, 1579-1946, 1587-1946, 1592-1959, 1636-1943, 1681-1925, 1690-2167, 1700-1928, 1706-1945, 1730-1999, 1732-2139, 1732-2720, 1736-2033, 1752-2209, 1758-2704, 1793-2213, 2061-2775, 2104-2781, 2717-3203, 2737-3197, 2740-3197, 2743-3197, 2962-3000, 2962-3016, 2962-3027, 2962-3052, 2962-3100, 2962-3116, 2962-3117, 2962-3118, 2962-3127, 2962-3130, 2962-3141, 2962-3175, 2962-3179, 2962-3186, 2962-3189, 2962-3196, 2962-3197, 2962-3198, 2962-3204, 2962-3206, 2962-3216, 2962-3345, 2962-3379, 2962-3406, 2962-3425, 2962-3489, 2962-3521, 2962-3644 |
| 70/7509919CB1/ 3092 | 1-83, 1-3068, 14-83, 38-563, 503-1013, 516-1302, 516-1349, 539-1195, 572-1078, 578-1077, 599-1090, 602-1215, 610-1204, 639-1215, 647-1007, 656-893, 680-964, 690-1122, 702-1199, 702-1224, 702-1231, 712-1605, 749-1002, 749-1010, 759-1011, 776-1343, 789-1252, 789-1389, 791-1280, 810-1376, 811-1435, 811-1469, 818-1109, 821-1121, 829-1256, 836-980, 842-1012, 844-1104, 870-1120, 883-1433, 908-1375, 910-1239, 910-1445, 914-1105, 932-1720, 949-1596, 954-1612, 957-1772, 966-1276, 968-1478, 970-1167, 975-1259, 977-1435, 986-1443, 995-1017, 997-1254, 1002-1262, 1002-1275, 1014-1163, 1027-1273, 1042-1280, 1042-1283, 1049-1513, 1060-1561, 1062-1315, 1063-1283, 1076-1548, 1084-1431, 1105-1333, 1106-1673, 1120-1849, 1126-1536, 1130-1570, 1140-1939, 1149-1844, 1152-1375, 1157-1424, 1158-1449, 1171-1412, 1177-1603, 1187-1725, 1193-1401, 1213-1654, 1229-1948, 1230-2074, 1244-1621, 1269-1726, 1282-1552, 1284-1534, 1287-1428, 1287-1563, 1287-1581, 1289-1723, 1289-1778, 1296-1532, 1304-1508, 1305-1803, 1306-1580, 1306-1871, 1308-1570, 1315-2020, 1316-1772, 1351-1543, 1355-1941, 1360-1606, 1360-1615, 1375-1848, 1383-2046, 1385-1653, 1387-1946, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 70 cont | 1413-1723, 1434-1625, 1437-2273, 1438-1687, 1441-1857, 1443-1699, 1445-1897, 1454-1723, 1454-1945, 1459-1647, 1459-1710, 1459-1722, 1459-1727, 1459-1750, 1460-1694, 1463-1736, 1464-2099, 1465-1776, 1467-1817, 1472-1998, 1472-2017, 1476-1686, 1478-1761, 1485-1745, 1503-1791, 1506-1729, 1507-1732, 1512-1670, 1519-1754, 1522-1742, 1522-1827, 1522-2059, 1526-1764, 1529-1759, 1530-1792, 1562-1838, 1581-2156, 1585-1867, 1592-2045, 1596-2080, 1597-2191, 1601-2262, 1607-2443, 1610-2086, 1620-2137, 1629-1916, 1631-2067, 1642-1952, 1648-2067, 1649-1874, 1653-1842, 1653-1855, 1655-1913, 1655-1927, 1655-2210, 1656-2171, 1659-1975, 1663-1902, 1663-1946, 1663-2240, 1676-1843, 1676-1867, 1678-1922, 1681-2307, 1682-1972, 1683-1916, 1683-1919, 1683-2001, 1689-2232, 1695-2235, 1697-1938, 1698-2149, 1701-1877, 1702-1949, 1708-2130, 1717-1997, 1717-2168, 1720-2337, 1725-2180, 1726-2337, 1730-1946, 1730-2046, 1741-2299, 1743-1878, 1745-2335, 1747-2246, 1752-1994, 1753-2227, 1773-2344, 1775-2027, 1775-2230, 1779-2063, 1790-2526, 1804-2352, 1812-2065, 1813-2001, 1815-1991, 1819-2008, 1830-2125, 1831-2089, 1838-2248, 1841-2095, 1844-2074, 1846-2133, 1850-2698, 1855-2038, 1868-2341, 1882-2114, 1889-2132, 1890-2174, 1900-2395, 1903-2332, 1905-2285, 1907-2491, |
| 70 cont | 1908-2182, 1911-2263, 1917-2384, 1917-2490, 1926-2231, 1930-2307, 1931-2115, 1931-2182, 1931-2428, 1934-2072, 1941-2443, 1943-2226, 1943-2248, 1945-2199, 1953-2235, 1953-2445, 1955-2227, 1965-2455, 1970-2609, 1985-2183, 1985-2241, 1986-2164, 1986-2300, 1987-2247, 1989-2085, 1992-2300, 1994-2493, 2003-2118, 2007-2118, 2007-2429, 2011-2392, 2012-2300, 2024-2474, 2037-2316, 2039-2299, 2052-2264, 2065-2307, 2066-2287, 2068-2322, 2068-2343, 2068-2372, 2068-2606, 2069-2499, 2084-2320, 2084-2352, 2084-2387, 2084-3007, 2088-2334, 2089-2332, 2090-2356, 2090-2365, 2097-2303, 2097-2447, 2102-2536, 2102-2786, 2102-2874, 2102-2955, 2102-3007, 2102-3033, 2103-2999, 2104-2991, 2124-2556, 2124-2594, 2133-3007, 2134-2358, 2136-2447, 2136-2684, 2145-3007, 2146-2544, 2146-3007, 2147-2312, 2147-2399, 2149-2480, 2150-3007, 2152-3007, 2154-2730, 2162-2430, 2163-3007, 2163-3050, 2164-3007, 2171-2426, 2172-2300, 2172-2371, 2172-2409, 2172-2429, 2172-2732, 2179-2647, 2179-2664, 2180-2399, 2180-2773, 2180-3006, 2183-2437, 2186-2574, 2190-3007, 2191-2502, 2191-2605, 2191-2621, 2192-2664, 2200-3007, 2204-3007, 2205-2447, 2205-2581, 2205-2586, 2205-2732, 2205-2802, 2207-2491, 2210-2622, 2210-3007, 2218-3007, |

Table 4

| Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 70 cont | 2226-3082, 2233-2977, 2234-2555, 2238-2930, 2240-2514, 2240-2543, 2244-3007, 2245-2352, 2246-2485, 2247-2545, 2248-2995, 2258-2469, 2263-2415, 2263-3082, 2265-2576, 2265-2720, 2265-3007, 2267-3007, 2268-2553, 2274-2490, 2275-2728, 2276-2780, 2281-2533, 2282-2555, 2287-2574, 2288-3010, 2293-3007, 2297-2580, 2297-2591, 2298-2530, 2298-3007, 2308-2447, 2308-2581, 2308-2721, 2308-2804, 2309-2565, 2309-2721, 2309-2732, 2309-3077, 2312-2564, 2313-2550, 2319-2720, 2331-3029, 2333-2635, 2341-3007, 2348-2576, 2348-2579, 2352-2602, 2352-2614, 2355-2602, 2358-2644, 2359-2775, 2359-3082, 2361-2712, 2361-2813, 2367-3054, 2367-3055, 2368-2616, 2368-3007, 2372-2429, 2384-2569, 2389-2624, 2391-2596, 2393-2801, 2396-2662, 2398-2659, 2398-2689, 2400-2625, 2404-2695, 2406-2674, 2410-2555, 2412-3007, 2423-2677, 2426-2678, 2426-2690, 2429-2693, 2429-2709, 2430-2581, 2430-2659, 2430-2670, 2430-2742, 2430-2801, 2430-2802, 2430-3029, 2431-2678, 2431-2709, 2431-3079, 2433-2698, 2439-2651, 2440-2708, 2441-3014, 2442-2697, 2442-3025, 2443-2677, 2443-2706, 2443-2892, 2444-2686, 2444-2998, 2448-2721, 2448-2742, 2454-3068, 2458-2736, 2459-3039, 2460-3020, 2471-2646, |
| 70 cont | 2471-2757, 2478-2737, 2480-2707, 2485-2937, 2486-2750, 2487-2757, 2487-2767, 2493-2799, 2496-2957, 2502-2765, 2502-3037, 2509-3007, 2510-2682, 2512-2756, 2523-2783, 2526-2793, 2527-3059, 2529-2742, 2533-2758, 2542-2749, 2545-2809, 2554-2809, 2555-2746, 2556-3039, 2569-3019, 2571-2836, 2573-3042, 2576-2834, 2576-2835, 2577-2832, 2581-2831, 2581-2862, 2582-2721, 2582-2806, 2582-2820, 2582-2832, 2601-2899, 2602-2822, 2603-2958, 2607-2877, 2607-3070, 2608-3077, 2610-2869, 2610-3076, 2611-2828, 2611-3070, 2613-3083, 2615-2946, 2617-3092, 2618-2971, 2619-3040, 2620-2845, 2621-2739, 2623-2979, 2623-3081, 2624-2779, 2624-2869, 2624-3054, 2625-3068, 2627-3042, 2629-3092, 2630-3077, 2630-3089, 2631-2889, 2634-2861, 2634-2869, 2634-3041, 2636-3092, 2637-2950, 2638-2887, 2640-2924, 2642-2847, 2642-3081, 2645-2904, 2645-3077, 2645-3091, 2646-3048, 2648-2976, 2650-3085, 2655-3068, 2656-2981, 2656-3078, 2657-2940, 2657-3040, 2657-3072, 2658-3007, 2658-3077, 2659-3079, 2660-3072, 2661-3042, 2662-3078, 2663-3084, 2664-2873, 2664-2942, 2664-3077, 2665-2864, 2665-2871, 2665-2897, 2665-2919, 2667-3078, 2669-3085, 2671-3066, 2671-3077, 2679-3077, 2682-2896, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 70 cont | 2682-2942, 2683-2987, 2683-3077, 2683-3080, 2691-2769, 2691-2930, 2693-3055, 2693-3092, 2698-3078, 2701-3090, 2703-3079, 2704-2951, 2706-3092, 2708-3086, 2709-3039, 2712-3015, 2716-3077, 2717-2987, 2718-3077, 2719-2997, 2720-2963, 2722-2979, 2723-3082, 2724-3078, 2730-2942, 2730-3008, 2730-3068, 2733-3007, 2743-3077, 2746-3077, 2756-3002, 2756-3006, 2756-3011, 2756-3078, 2758-3010, 2758-3077, 2759-3081, 2768-3042, 2768-3077, 2769-3077, 2786-3058, 2786-3092, 2790-3026, 2790-3055, 2792-3072, 2793-3062, 2794-2971, 2796-3042, 2797-3067, 2798-2964, 2803-3067, 2807-3037, 2807-3069, 2807-3077, 2811-3082, 2814-3091, 2818-3077, 2824-3072, 2832-3077, 2839-3082, 2845-3077, 2845-3092, 2851-3079, 2859-3092, 2861-3081, 2861-3092, 2865-3090, 2865-3092, 2867-3061, 2868-3085, 2874-3092, 2878-3088, 2880-3077, 2880-3086, 2880-3092, 2885-3092, 2891-3092, 2915-3092, 2917-3078 |
| 71/7510758CB1/ 3257 | 1-449, 31-343, 34-448, 390-1014, 442-1021, 455-924, 455-1167, 455-3257, 506-1049, 564-1129, 806-1615, 806-1619, 807-1564, 818-1604, 836-1522, 939-1539, 942-1800, 978-1297, 1066-1548, 1067-1548, 1090-1408, 1116-1548, 1187-1513, 1187-1537, 1187-1539, 1187-1542, 1191-1503, 1191-1539, 1226-1976, 1259-1548, 1285-1548, 1343-1548, 1404-1539, 1475-2276, 1494-2072, 1544-2288, 1544-2357, 1547-2233, 1547-2276, 1549-2041, 1549-2199, 1549-2205, 1549-2233, 1549-2254, 1550-2186, 1557-2357, 1558-2357, 1562-2408, 1562-2460, 1562-2465, 1565-2048, 1567-2237, 1589-2357, 1624-2357, 1632-2357, 1655-2357, 1667-2357, 1688-2357, 1696-2357, 1713-2357, 1715-2622, 1719-2357, 1740-2357, 1756-2357, 1757-2618, 1761-2357, 1762-2626, 1766-2357, 1805-2247, 1838-2357, 1856-2357, 1870-2408, 1870-2518, 1876-2357, 1895-2670, 1940-2358, 2346-3088, 2346-3169, 2346-3183, 2346-3200, 2376-3163, 2399-3248, 2401-3248 |
| 72/7510063CB1/ 3360 | 1-611, 26-3360, 104-535, 181-737, 521-894, 842-1292, 1017-1654, 1019-1647, 1088-1189, 1585-2181, 1690-2288, 1692-2341, 1695-2349, 1695-2350, 1695-2368, 1695-2387, 1702-2011, 1728-2394, 1730-2338, 1732-2255, 1732-2316, 1742-2322, 1745-1991, 1812-2446, 1816-2338, 1851-2273, 1854-2431, 1855-2426, 1858-2156, 1866-2455, 1950-2536, 1959-2338, 1962-2478, 1962-2651, 1973-2910, 1977-2247, 1981-2448, 1986-2651, 1987-2228, 1996-2651, 2005-2651, 2014-2651, 2017-2632, 2032-2338, 2032-2489, 2033-2651, 2037-2644, 2038-2651, 2041-2627, 2104-2818, 2123-2753, 2163-2906, 2165-2377, 2165-2615, 2233-2515, 2250-2838, 2283-2544, 2288-2872, 2291-2734, 2297-2534, 2353-2889, 2362-2636, 2370-2880, 2378-2950, 2379-2945, 2398-2948, 2408-2607, 2414-2701, 2465-2838, 2466-2904, 2483-2897, 2570-2789, 2586-2971, 2638-2913, 2646-2880, 2762-2854, 2762-2882, 2762-2984 |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 73/7510135CBI/ 6871 | 1-78, 1-6855, 20-75, 20-78, 191-800, 315-568, 363-1004, 1038-1309, 1038-1480, 1073-1292, 1113-1672, 1502-1759, 1653-2388, 1698-2253, 1745-2288, 1804-2327, 1807-2245, 1820-2245, 1820-2305, 1855-2431, 1989-2273, 2115-2232, 2138-2729, 2229-2463, 2255-2684, 2325-2837, 2349-2800, 2379-2682, 2407-3050, 2419-2441, 2551-2796, 2608-2761, 2609-2866, 2632-2866, 2671-2905, 2672-2907, 2694-3285, 2727-3321, 2737-3285, 2740-3285, 2796-3000, 2815-3114, 2935-3258, 2938-3595, 2988-3243, 3005-3595, 3055-3312, 3096-3491, 3105-3383, 3137-3684, 3197-3436, 3306-3586, 3311-3643, 3429-3654, 3573-3867, 3608-4354, 3634-3770, 3635-3985, 3649-4188, 3668-4164, 3729-4237, 3749-4171, 3778-4195, 3780-4042, 3832-4388, 3840-4089, 3880-4446, 3981-4178, 4073-4353, 4103-4558, 4140-4410, 4172-4635, 4177-4436, 4177-4697, 4180-4468, 4184-4697, 4187-4784, 4200-4427, 4257-4659, 4258-4608, 4266-4556, 4296-4554, 4301-4829, 4303-4458, 4309-4579, 4309-4606, 4434-4969, 4472-4710, 4478-4566, 4500-4809, 4504-4635, 4505-4723, 4505-4955, 4526-4955, 4530-4801, 4530-5170, 4548-4823, 4562-4839, 4590-5190, 4594-4837, 4599-5148, 4613-4880, 4616-4850, 4622-4892, 4622-4905, 4709-4993, 4740-4914, 4802-5081, 4823-5091, |
| 73 cont | 4852-5151, 4864-5121, 4869-5043, 4874-4984, 4923-5163, 4928-5185, 4930-5164, 4932-5732, 4954-5503, 4962-5189, 4978-5368, 4980-5160, 5038-5301, 5038-5674, 5044-5612, 5060-5509, 5061-5359, 5061-5495, 5062-5294, 5073-5330, 5091-5331, 5094-5572, 5099-5296, 5099-5368, 5100-5384, 5102-5290, 5125-5377, 5131-5383, 5150-5357, 5161-5543, 5169-5318, 5178-5462, 5179-5416, 5192-5403, 5194-5437, 5198-5441, 5203-5513, 5223-5497, 5246-5500, 5253-5394, 5254-5522, 5260-5540, 5263-5584, 5264-5493, 5289-5506, 5289-5524, 5289-5563, 5318-5548, 5323-5579, 5341-5578, 5347-5597, 5347-5958, 5368-5649, 5375-5675, 5387-5627, 5387-5635, 5387-5878, 5402-5690, 5415-5648, 5415-5674, 5415-5957, 5454-5645, 5454-5722, 5466-5601, 5470-5719, 5482-5732, 5492-5745, 5496-5720, 5496-5738, 5497-5786, 5497-5799, 5510-6111, 5527-6181, 5528-5641, 5528-5759, 5538-5799, 5538-5806, 5544-5885, 5545-5826, 5545-5828, 5546-5793, 5549-6076, 5557-5837, 5576-5881, 5580-5855, 5588-5884, 5597-5878, 5597-5891, 5606-5895, 5620-5945, 5621-5869, 5640-6169, 5643-5901, 5651-5897, 5652-5956, 5663-5810, 5663-6359, 5664-6289, 5668-5883, 5668-5918, 5671-5906, 5671-5945, 5671-6196, 5671-6223, 5672-5872, 5680-6223, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 73 cont | 5694-5921, 5696-5928, 5700-5967, 5704-5960, 5705-5988, 5710-5988, 5713-5994, 5729-5986, 5731-6015, 5732-5997, 5732-6001, 5732-6256, 5751-6021, 5765-6370, 5767-6030, 5802-6057, 5802-6063, 5802-6071, 5809-6273, 5838-6102, 5841-6194, 5849-6486, 5851-6111, 5862-6127, 5864-6150, 5867-6112, 5868-6165, 5868-6291, 5869-6118, 5870-6115, 5873-6118, 5882-6103, 5883-6130, 5883-6144, 5883-6170, 5884-6413, 5891-6149, 5894-6148, 5894-6159, 5894-6389, 5895-6156, 5902-6162, 5904-6102, 5904-6116, 5906-6163, 5907-6043, 5909-6197, 5909-6198, 5915-6150, 5919-6482, 5933-6170, 5936-6233, 5943-6136, 5943-6234, 5948-6460, 5956-6209, 5966-6242, 5969-6200, 5971-6269, 5976-6256, 5978-6223, 5978-6252, 5982-6284, 5983-6489, 5993-6443, 6002-6413, 6006-6253, 6006-6295, 6006-6382, 6007-6270, 6008-6301, 6008-6309, 6009-6269, 6009-6276, 6010-6255, 6017-6774, 6020-6270, 6020-6281, 6020-6324, 6021-6308, 6023-6246, 6023-6503, 6024-6321, 6028-6280, 6028-6284, 6028-6442, 6029-6331, 6041-6287, 6042-6328, 6043-6268, 6045-6229, 6046-6601, 6047-6314, 6048-6313, 6049-6331, 6059-6348, 6060-6329, 6063-6293, 6066-6297, 6066-6633, 6071-6317, 6076-6320, 6076-6581, 6077-6315, 6085-6358, 6086-6368, |
| 73 cont | 6086-6378, 6090-6334, 6091-6345, 6095-6350, 6095-6372, 6095-6387, 6095-6395, 6105-6483, 6110-6355, 6110-6741, 6114-6360, 6114-6366, 6114-6380, 6142-6398, 6143-6409, 6154-6338, 6162-6437, 6179-6475, 6187-6446, 6190-6490, 6190-6789, 6191-6429, 6209-6425, 6209-6686, 6209-6778, 6209-6795, 6209-6817, 6214-6418, 6223-6359, 6223-6455, 6227-6527, 6231-6840, 6241-6508, 6241-6526, 6243-6811, 6246-6502, 6248-6498, 6248-6511, 6248-6528, 6252-6697, 6255-6857, 6274-6529, 6283-6840, 6284-6544, 6284-6568, 6286-6813, 6287-6840, 6289-6829, 6296-6871, 6297-6544, 6297-6818, 6298-6818, 6300-6597, 6306-6562, 6307-6570, 6310-6567, 6316-6563, 6316-6566, 6319-6533, 6320-6871, 6322-6752, 6326-6871, 6332-6575, 6334-6871, 6340-6634, 6352-6607, 6353-6597, 6354-6871, 6356-6609, 6357-6625, 6357-6629, 6362-6599, 6362-6600, 6362-6621, 6365-6822, 6367-6842, 6373-6628, 6374-6627, 6375-6652, 6376-6680, 6377-6648, 6380-6640, 6386-6817, 6389-6671, 6392-6657, 6392-6871, 6393-6829, 6393-6830, 6393-6832, 6393-6841, 6393-6856, 6394-6832, 6394-6860, 6395-6674, 6395-6856, 6395-6859, 6396-6864, 6399-6673, 6399-6857, 6406-6638, 6406-6653, 6406-6710, 6406-6861, 6408-6656, 6408-6669, 6408-6857, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 73 cont | 6413-6691, 6413-6784, 6416-6711, 6419-6609, 6421-6864, 6424-6841, 6424-6857, 6425-6856, 6428-6857, 6429-6674, 6429-6710, 6429-6715, 6430-6660, 6434-6856, 6435-6696, 6440-6860, 6444-6864, 6445-6717, 6449-6856, 6449-6871, 6450-6681, 6450-6857, 6451-6857, 6452-6856, 6452-6862, 6454-6856, 6455-6732, 6457-6856, 6458-6858, 6459-6862, 6459-6865, 6460-6856, 6461-6857, 6465-6865, 6466-6729, 6467-6856, 6471-6856, 6472-6857, 6473-6857, 6477-6614, 6477-6857, 6482-6850, 6485-6871, 6497-6856, 6498-6750, 6498-6818, 6508-6856, 6512-6789, 6512-6856, 6512-6857, 6520-6856, 6522-6773, 6522-6856, 6523-6862, 6525-6724, 6525-6772, 6526-6746, 6545-6856, 6546-6856, 6555-6858, 6558-6856, 6563-6855, 6565-6832, 6567-6821, 6570-6857, 6576-6856, 6577-6839, 6588-6843, 6589-6843, 6589-6856, 6593-6734, 6597-6846, 6609-6855, 6616-6723, 6618-6857, 6618-6871, 6620-6754, 6620-6870, 6634-6865, 6639-6856, 6643-6871, 6649-6833, 6657-6851, 6657-6871, 6665-6856, 6683-6856, 6688-6871, 6701-6860, 6704-6871, 6709-6862, 6713-6871, 6714-6871, 6726-6860, 6750-6857, 6750-6871, 6757-6871, 6784-6871, 6789-6869, 6792-6871 |
| 74/7505011CB1/ 6696 | 1-6633, 187-786, 458-920, 525-782, 562-1110, 613-1086, 616-813, 621-1205, 734-785, 737-803, 740-804, 749-1004, 749-1212, 749-1225, 825-1060, 855-1093, 977-1553, 992-1561, 1022-1085, 1022-1086, 1046-1698, 1086-1346, 1126-1345, 1185-1292, 1185-1510, 1185-1719, 1185-1737, 1204-1292, 1204-1523, 1208-1562, 1266-1292, 1281-1461, 1281-1499, 1282-1460, 1282-1475, 1282-1485, 1282-1486, 1317-1527, 1339-2018, 1371-1478, 1371-1709, 1371-1923, 1375-1737, 1375-2022, 1378-1737, 1388-1436, 1388-1439, 1390-1478, 1394-1457, 1394-1709, 1426-1902, 1431-1925, 1444-1499, 1507-2008, 1537-1592, 1544-2224, 1561-1897, 1561-1923, 1564-1610, 1653-1711, 1653-1737, 1654-1737, 1654-1847, 1654-1857, 1655-1857, 1660-1832, 1663-1737, 1666-2243, 1668-1910, 1674-1757, 1712-1957, 1712-1966, 1723-1778, 1723-1785, 1730-1974, 1731-2393, 1758-2017, 1782-2399, 1816-1871, 1835-2420, 1838-2296, 1839-1923, 1895-2636, 1946-2417, 2035-2756, 2058-2715, 2070-2485, 2097-2722, 2113-2745, 2143-2367, 2195-2751, 2224-2468, 2234-2710, 2281-2548, 2395-2932, 2407-2986, 2501-3096, 2520-2951, 2575-2878, 2647-2942, 2691-3173, 2703-3228, 2740-2860, 2807-2829, 2826-3254, 2870-3163, 2921-3168, 2966-3299, 3031-3316, 3034-3322, 3039-3333, 3048-3272, 3093-3318, 3164-3786, 3216-3466, 3344-3577, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|--|
| 74 cont | 3344-3579, 3359-3758, 3364-3764, 3369-3631, 3371-3407, 3371-3439, 3371-3458, 3371-3514, 3371-3527, 3371-3567, 3371-3616, 3379-3461, 3379-3494, 3379-3738, 3379-4101, 3381-3684, 3397-3616, 3413-3616, 3432-3686, 3434-3616, 3435-3616, 3465-4130, 3468-3770, 3516-3616, 3516-4139, 3533-3616, 3535-3616, 3572-3616, 3576-3616, 3592-4158, 3601-4211, 3633-4277, 3652-3734, 3652-3764, 3652-3767, 3652-4013, 3652-4374, 3654-3957, 3670-3764, 3700-3764, 3710-3764, 3837-4346, 3840-4031, 3840-4422, 3857-4394, 3876-4134, 3905-4310, 3919-4011, 3919-4325, 3978-4258, 3982-4226, 4013-4310, 4014-4309, 4048-4333, 4190-4285, 4190-4386, 4190-4435, 4198-4310, 4200-4435, 4216-4528, 4232-4853, 4253-4522, 4254-4498, 4284-4435, 4288-4435, 4335-4435, 4335-4926, 4349-4859, 4354-4877, 4391-4627, 4395-4729, 4414-4435, 4416-4706, 4435-4876, 4453-4700, 4453-4949, 4471-4857, 4594-5243, 4598-5130, 4637-5309, 4712-5185, 4718-5001, 4725-4826, 4743-5337, 4755-5351, 4757-5145, 4805-5197, 4827-5267, 4857-5457, 4857-5591, 4864-5455, 4913-5156, 4930-5202, 4937-5230, 4937-5435, 4937-5441, 4937-5479, 4940-5454, 4948-5197, 4984-5423, 4986-5367, 5002-5619, 5005-5267, 5010-5254, 5013-5213, 5013-5269, 5023-5651, |
| 74 cont | 5030-5572, 5030-5626, 5038-5300, 5038-5315, 5038-5466, 5046-5667, 5049-5322, 5049-5647, 5054-5270, 5062-5305, 5065-5488, 5074-5749, 5078-5321, 5099-5657, 5108-5362, 5124-5326, 5135-5343, 5136-5346, 5136-5370, 5136-5393, 5145-5660, 5154-5628, 5195-5437, 5197-5826, 5206-5380, 5208-5745, 5210-5727, 5210-5777, 5212-5426, 5212-5803, 5214-5748, 5217-5795, 5227-5535, 5234-5912, 5240-5579, 5246-5506, 5249-5661, 5255-5526, 5266-5760, 5269-5836, 5278-5526, 5278-5531, 5281-5574, 5288-5621, 5296-5527, 5298-5879, 5304-5671, 5306-5890, 5310-5590, 5312-5578, 5312-5584, 5315-5472, 5320-5825, 5337-5599, 5344-5632, 5362-5619, 5366-5585, 5370-5625, 5371-5656, 5372-5642, 5381-5693, 5389-5653, 5392-5758, 5395-5660, 5395-5670, 5397-5922, 5398-5976, 5412-5985, 5413-5693, 5413-5865, 5417-5861, 5439-5674, 5444-5753, 5444-5959, 5455-5751, 5457-5693, 5460-5970, 5471-6075, 5488-6091, 5510-5765, 5510-5776, 5517-5631, 5517-5760, 5517-5770, 5534-6119, 5538-6185, 5541-6154, 5543-5778, 5543-5780, 5549-6046, 5553-5694, 5561-5665, 5561-5698, 5571-6148, 5580-6046, 5582-6128, 5582-6175, 5583-6216, 5587-5794, 5596-5792, 5601-5805, 5608-5708, 5608-5931, 5608-6476, 5609-5893, 5610-5882, 5611-5867, |

Table 4

| Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length | Sequence Fragments |
|--|---|
| 74 cont | 5631-5828, 5632-5887, 5632-6180, 5632-6215, 5634-5785, 5634-5898, 5634-5918, 5639-5929, 5641-6091, 5650-6154, 5662-5938, 5662-5994, 5662-6017, 5662-6067, 5664-6341, 5689-5902, 5692-5993, 5709-6012, 5714-5961, 5717-5996, 5720-6282, 5722-5968, 5724-5982, 5724-5999, 5728-5937, 5728-6056, 5728-6456, 5730-5968, 5730-6015, 5740-6034, 5750-6467, 5752-5999, 5752-6036, 5757-6018, 5761-6250, 5764-6364, 5766-6005, 5776-5954, 5776-6368, 5785-5907, 5785-5993, 5789-5999, 5799-6045, 5799-6444, 5812-6049, 5816-6058, 5820-6044, 5820-6059, 5820-6283, 5822-6420, 5824-6060, 5826-6265, 5826-6359, 5827-6070, 5828-6352, 5828-6396, 5830-6174, 5832-6066, 5832-6077, 5832-6098, 5832-6111, 5833-6490, 5840-6102, 5842-6087, 5842-6092, 5847-6390, 5849-6119, 5850-6125, 5853-6474, 5854-6077, 5858-6390, 5863-6186, 5866-6462, 5867-6530, 5868-6507, 5869-6072, 5870-6103, 5870-6135, 5876-6134, 5876-6140, 5876-6398, 5882-6137, 5884-6383, 5890-6158, 5901-6120, 5902-6144, 5903-6530, 5904-6221, 5909-6453, 5914-6191, 5920-6168, 5921-6150, 5921-6211, 5924-6151, 5924-6172, 5924-6225, 5924-6440, 5926-6435, 5928-6212, 5931-6061, 5942-6610, 5943-6262, 5959-6208, 5966-6219, |
| 74 cont | 5977-6430, 5978-6247, 5985-6221, 5991-6435, 5995-6672, 5998-6239, 5999-6306, 6001-6269, 6006-6259, 6015-6639, 6017-6280, 6034-6511, 6044-6344, 6058-6639, 6059-6313, 6060-6332, 6060-6363, 6070-6220, 6078-6338, 6078-6551, 6080-6334, 6082-6381, 6084-6333, 6094-6351, 6101-6304, 6103-6337, 6103-6389, 6104-6627, 6105-6366, 6118-6675, 6126-6432, 6131-6594, 6132-6408, 6139-6390, 6140-6465, 6150-6431, 6157-6594, 6168-6380, 6168-6389, 6168-6405, 6176-6406, 6180-6512, 6184-6468, 6185-6403, 6185-6630, 6195-6448, 6201-6479, 6207-6419, 6209-6421, 6211-6480, 6212-6466, 6217-6564, 6218-6492, 6227-6452, 6227-6486, 6231-6539, 6242-6515, 6242-6538, 6270-6527, 6270-6532, 6276-6542, 6282-6523, 6284-6414, 6290-6568, 6294-6540, 6294-6560, 6294-6568, 6297-6563, 6306-6564, 6308-6602, 6330-6603, 6333-6575, 6342-6557, 6342-6564, 6342-6625, 6343-6599, 6349-6618, 6355-6603, 6361-6615, 6363-6652, 6370-6658, 6373-6696, 6375-6556, 6379-6512, 6379-6518, 6379-6583, 6379-6626, 6385-6609, 6404-6587, 6411-6640, 6412-6673, 6420-6644, 6437-6603, 6475-6632 |

Table 5

| Polynucleotide SEQ ID NO: | Incyte Project ID: | Representative Library |
|------------------------------|--------------------|------------------------|
| 38 | 7504868CB1 | COLNFEC01 |
| 39 | 7504930CB1 | BRAIFER05 |
| 40 | 6610456CB1 | OVARDIR01 |
| 41 | 7503573CB1 | SINTFER02 |
| 42 | 7505057CB1 | UCMCL5T01 |
| 43 | 90116002CB1 | DUODNOT02 |
| 44 | 039283CB1 | BLADTUT05 |
| 45 | 7505082CB1 | LIVRNON08 |
| 46 | 7505139CB1 | LUNGNOT23 |
| 47 | 7505234CB1 | LUNPTUT02 |
| 48 | 7500227CB1 | BRSTTUT03 |
| 49 | 7503676CB1 | FTUBTUE01 |
| 50 | 7503606CB1 | BRAHTDR03 |
| 51 | 7500216CB1 | SINTNOR01 |
| 52 | 7099880CB1 | BRAHTDR04 |
| 53 | 871513CB1 | MIXDUNB01 |
| 54 | 8057640CB1 | EPIMNON05 |
| 55 | 7505913CB1 | NOSETUE01 |
| 56 | 7510292CB1 | PLACNOR01 |
| 57 | 7504669CB1 | BRAUNOR01 |
| 58 | 7509266CB1 | KIDPTDE01 |
| 59 | 7509288CB1 | LUNGDIS03 |
| 60 | 7510212CB1 | URETTUE01 |
| 61 | 7510504CB1 | NERDTDN03 |
| 62 | 7510587CB1 | OVARTUT10 |
| 63 | 7510684CB1 | BRAITUT02 |
| 64 | 7510697CB1 | PLACNOR01 |
| 65 | 7761337CB1 | PENITUT01 |
| 66 | 7503666CB1 | SINTNOR01 |
| 67 | 7503668CB1 | SINTNOR01 |
| 68 | 7503672CB1 | SINTNOR01 |
| 69 | 6039650CB1 | SKINDIA01 |
| 70 | 7509919CB1 | MMLR2DT01 |
| 71 | 7510758CB1 | SYNWDIT01 |
| 72 | 7510063CB1 | PROTDNV25 |
| 73 | 7510135CB1 | LUNGNOT23 |
| 74 | 7505011CB1 | BONEUNR01 |

Table 6

| Library | Vector | Library Description |
|-----------|-------------|--|
| BLADTUT05 | pINCY | Library was constructed using RNA isolated from bladder tumor tissue removed from a 66-year-old Caucasian male during a radical prostatectomy, radical cystectomy, and urinary diversion. Pathology indicated grade 3 transitional cell carcinoma on the anterior wall of the bladder. Patient history included lung neoplasm and tobacco abuse in remission. Family history included malignant breast neoplasm, tuberculosis, cerebrovascular disease, atherosclerotic coronary artery disease, and lung cancer. |
| BRAIFER05 | pINCY | Library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. |
| COLNFEC01 | pINCY | This large size-fractionated library was constructed using RNA isolated from colon tissue removed from a Caucasian male fetus who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation. |
| DUODNOT02 | pINCY | Library was constructed using RNA isolated from duodenal tissue of an 8-year-old Caucasian female, who died from head trauma. Serology was positive for cytomegalovirus (CMV). |
| OVARDIR01 | PCDNA2.1 | This random primed library was constructed using RNA isolated from right ovary tissue removed from a 45-year-old Caucasian female during total abdominal hysterectomy, bilateral salpingo-oophorectomy, vaginal suspension and fixation, and incidental appendectomy. Pathology indicated stromal hyperthecosis of the right and left ovaries. Pathology for the matched tumor tissue indicated a dermoid cyst (benign cystic teratoma) in the left ovary. Multiple (3) intramural leiomyomata were identified. The cervix showed squamous metaplasia. Patient history included metrorrhagia, female stress incontinence, alopecia, depressive disorder, pneumonia, normal delivery, and deficiency anemia. Family history included benign hypertension, atherosclerotic coronary artery disease, hyperlipidemia, and primary tuberculous complex. |
| SINTFER02 | pINCY | This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise. |
| UCMCL5T01 | PBLUESCRIPT | Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of 12 individuals. The cells were cultured for 12 days with IL-5 before RNA was obtained from the pooled lysates. |
| BRSTTUT03 | PSPORT1 | Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes. |

Table 6

| Library | Vector | Library Description |
|-----------|--------|---|
| FTUBTUE01 | pINCY | This 5' biased random primed library was constructed using RNA isolated from right fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma of the right fallopian tube, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. A metastatic endometrioid and serous adenocarcinoma was present in the cul-de-sac tumor. The patient presented with a pelvic mass and ascites. Patient history included medullary carcinoma of the thyroid and myocardial infarction. Patient medications included Nitro-Dur, Lescol, Lasix and Cardizem. |
| LIVRNON08 | pINCY | This normalized library was constructed from 5.7 million independent clones from a pooled liver tissue library. Starting RNA was made from pooled liver tissue removed from a 4-year-old Hispanic male who died from anoxia and a 16 week female fetus who died after 16-weeks gestation from anencephaly. Serologies were positive for cytomegalovirus in the 4-year-old. Patient history included asthma in the 4-year-old. Family history included taking daily prenatal vitamins and mitral valve prolapse in the mother of the fetus. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. |
| LUNGNOT23 | pINCY | Library was constructed using RNA isolated from left lobe lung tissue removed from a 58-year-old Caucasian male. Pathology for the associated tumor tissue indicated metastatic grade 3 (of 4) osteosarcoma. Patient history included soft tissue cancer, secondary cancer of the lung, prostate cancer, and an acute duodenal ulcer with hemorrhage. Family history included prostate cancer, breast cancer, and acute leukemia. |
| LUNPTUT02 | pINCY | Library was constructed using RNA isolated from pleura tumor tissue removed from a 55-year-old Caucasian female during complete pneumonectomy. Pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, forming a bossellated mass replacing the right lower lobe and a portion of the middle lobe. The tumor involved the adjacent parietal pleura and pericardium. Multiple nodules comprising the tumor show near total necrosis. The right upper lobe was atelectic but uninvolved by tumor. Microsections of cellular nodules show brisk mitotic activity. The pericardium shows direct involvement but its margins were tumor free. Smooth muscle actin was positive. Estrogen receptor was negative and progesterone receptor was positive. Patient history included shortness of breath, peptic ulcer disease, lung cancer, uterine cancer, normal delivery, tobacco abuse, and deficiency anemia. Previous surgeries included endoscopic excision of a lung lesion. Family history included atherosclerotic coronary artery disease, breast cancer, type II diabetes, and multiple sclerosis. |

Table 6

| Library | Vector | Library Description |
|-----------|----------|--|
| BRAHTDR03 | PCDNA2.1 | This random primed library was constructed using RNA isolated from archaetox, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver. |
| BRAHTDR04 | PCDNA2.1 | This random primed library was constructed using RNA isolated from archaetox, anterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver. |
| MIXDUNB01 | pINCY | Library was constructed using RNA isolated from myometrium removed from a 41-year-old Caucasian female (A) during vaginal hysterectomy with a dilatation and curettage and untreated smooth muscle cells removed from the renal vein of a 57-year-old Caucasian male. Pathology for donor A indicated the myometrium and cervix were unremarkable. The endometrium was secretory and contained fragments of endometrial polyps. Benign endo- and ectocervical mucosa were identified in the endocervix. Pathology for the associated tumor tissue indicated uterine leiomyoma. Medical history included an unspecified menstrual disorder, ventral hernia, normal delivery, a benign ovarian neoplasm, and tobacco abuse in donor A. Previous surgeries included a bilateral destruction of fallopian tubes, removal of a solitary ovary, and an exploratory laparotomy in donor A. Medications included ferrous sulfate in donor A. |
| SINTNOR01 | PCDNA2.1 | This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity. |

Table 6

| Library | Vector | Library Description |
|-----------|----------|---|
| EPIMNON05 | pINCY | This normalized mammary epithelial cell tissue library was constructed from 3.28 million independent clones from an epithelial cell tissue library. Starting RNA was made from untreated mammary epithelial cell tissue removed from a 21-year-old female. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 -hours/round) reannealing hybridization was used. |
| NOSETUE01 | PCDNA2.1 | This 5' biased random primed library was constructed using RNA isolated from nasal and cribriform tumor tissue removed from a 45-year-old Caucasian male during total face osteotomy with reconstruction, rhinotomy and craniotomy. Pathology indicated olfactory neuroblastoma in the nasal cavity and cribriform region. The patient presented with cancer of the head, face and neck, and epistaxis. Patient history included extrinsic asthma, cancer of the head, face and neck, and epistaxis. Previous surgeries included total face osteotomy with reconstruction. Patient medications included Biaxin, Atessalon, and Valium. The patient received radiation treatments. Family history included chronic sinusitis in the mother and type II diabetes in the father. |
| BRAITUT02 | PSPORT1 | Library was constructed using RNA isolated from brain tumor tissue removed from the frontal lobe of a 58-year-old Caucasian male during excision of a cerebral meningeal lesion. Pathology indicated a grade 2 metastatic hypernephroma. Patient history included a grade 2 renal cell carcinoma, insomnia, and chronic airway obstruction. Family history included a malignant neoplasm of the kidney. |
| BRAUNOR01 | pINCY | This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages. |

Table 6

| Library | Vector | Library Description |
|-----------|----------|---|
| KIDPTDE01 | PCDNA2.1 | This 5' biased random primed library was constructed using RNA isolated from diseased left kidney tissue removed from a 63-year-old Caucasian male during nephroureterectomy, radical prostatectomy, urinary diversion to bowel, and regional lymph node excision. Pathology indicated minimal chronic interstitial nephritis. Pathology for the associated tumor tissue indicated grade 4 transitional cell carcinoma with squamous differentiation involving the lateral bladder wall and extending superficially into the left trigone which shows deep muscle invasion. Tumor encases the proximal left ureter. The patient presented with unspecified abdominal and pelvic symptoms, hematuria, and hyperplasia of the prostate. Patient history included atrial fibrillation and tobacco abuse. The patient was not taking any medications. Family history included cerebrovascular accident and atherosclerotic coronary artery disease in the mother and acute myocardial infarction and atherosclerotic coronary artery disease in the father. |
| LUNGDIS03 | pINCY | Library was constructed using diseased lung tissue. 0.76 million clones from a diseased lung tissue library were subjected to two rounds of subtraction hybridization with 5.1 million clones from a normal lung tissue library. The starting library for subtraction was constructed using polyA RNA isolated from diseased lung tissue. Patient history included idiopathic pulmonary disease. Subtractive hybridization conditions were based on the methodologies of Swaroop et al. (1991) Nucleic Acids Res. 19:1954; and Bonaldo et al. Genome Res. (1996) 6:791. |
| NERDTDN03 | pINCY | This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone), hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used. |

Table 6

| Library | Vector | Library Description |
|-----------|----------|--|
| OVARTUT10 | pINCY | Library was constructed using RNA isolated from ovarian tumor tissue removed from the left ovary of a 58-year-old Caucasian female during a total abdominal hysterectomy, removal of a solitary ovary, and repair of inguinal hernia. Pathology indicated a metastatic grade 3 adenocarcinoma of colonic origin, forming a partially cystic and necrotic tumor mass in the left ovary, and an adenocarcinoma of colonic origin, forming a nodule in the left mesovarium. A single intramural leiomyoma was identified in the myometrium. The cervix showed mild chronic cystic cervicitis. Patient history included benign hypertension, follicular cyst of the ovary, colon cancer, benign colon neoplasm, and osteoarthritis. Family history included emphysema, myocardial infarction, atherosclerotic coronary artery disease, benign hypertension, and hyperlipidemia. |
| PENTUT01 | pINCY | Library was constructed using RNA isolated from tumor tissue removed from the penis of a 64-year-old Caucasian male during penile amputation. Pathology indicated a fungating invasive grade 4 squamous cell carcinoma involving the inner wall of the foreskin and extending onto the glans penis. Patient history included benign neoplasm of the large bowel, atherosclerotic coronary artery disease, angina pectoris, gout, and obesity. Family history included malignant pharyngeal neoplasm, chronic lymphocytic leukemia, and chronic liver disease. |
| PLACNOR01 | PCDNA2.1 | This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from placental tissue removed from a Caucasian fetus (donor A), who died after 16 weeks' gestation from fetal demise and hydrocephalus and from placental tissue removed from a Caucasian male fetus (donor B), who died after 18 weeks' gestation from fetal demise. Patient history for donor A included umbilical cord wrapped around the head (3 times) and the shoulders (1 time). Serology was positive for anti-CMV and remaining serologies were negative. Family history included multiple pregnancies and live births, and an abortion in the mother. Serology was negative for donor B. |
| URETTUE01 | PCDNA2.1 | This 5' biased random primed library was constructed using RNA isolated from ureter tumor tissue removed from a 64-year-old Caucasian male during closed bladder biopsy, radical cystectomy, radical prostatectomy, and formation of a cutaneous ureterostomy. Pathology indicated in situ and superficially invasive transitional cell carcinoma presenting as 2 separate papillary lesions, one located 7.5 cm from the ureter margin, and the other in the right proximal ureter extending into the renal pelvis. The tumor invaded just into the submucosal tissue. The ureter margin was involved by focal in situ transitional cell carcinoma. The patient presented with carcinoma in situ of the bladder, malignant neoplasm of the ureter, and secondary malignant kidney neoplasm. Patient history included malignant bladder neoplasm, psoriasis, chronic airway obstruction, testicular hypofunction, and tobacco abuse. Previous surgeries included appendectomy and transurethral destruction of bladder lesion. Patient medications included naproxen, Atrovent, albuterol, and an unspecified psoriasis cream. Family history included malignant stomach neoplasm in the father and malignant bladder neoplasm |

Table 6

| Library | Vector | Library Description |
|-----------|--------------|---|
| BONEUNR01 | PCDNA2.1 | This random primed library was constructed using pooled cDNA from two different donors. cDNA was generated using mRNA isolated from an untreated MG-63 cell line derived from an osteosarcoma tumor removed from a 14-year-old Caucasian male (donor A) and using mRNA isolated from sacral bone tumor tissue removed from an 18-year-old Caucasian female (donor B) during an exploratory laparotomy and soft tissue excision. Pathology indicated giant cell tumor of the sacrum in donor B. Donor B's history included pelvic joint pain, constipation, urinary incontinence, unspecified abdominal/pelvic symptoms, and a pelvic soft tissue malignant neoplasm. Family history included prostate cancer in donor B. |
| MMLR2DT01 | PSPORT1 | Library was constructed using RNA isolated from plastic adherent mononuclear cells isolated from buffy coat units obtained from unrelated male and female donors. |
| PROTDNV25 | pCR2-TOPO-TA | Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled small intestine tissue removed from a Caucasian male fetus (donor A) who died at 23 weeks' gestation from premature birth; from lung tissue removed from a Caucasian male fetus (donor B) who died from fetal demise; from pleura tumor tissue removed from a 55-year-old Caucasian female (donor C) during a complete pneumonectomy; from frontal/parietal brain tumor tissue removed from a 2-year-old Caucasian female (donor D) during excision of cerebral meningeal lesion; from liver tumor tissue removed from a 72-year-old Caucasian male (donor E) during partial hepatectomy; from pooled fetal brain tissue removed from a Caucasian male fetus (donor F) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor G), who died at 23 weeks' gestation from premature birth; from pooled fetal kidney tissue removed from 59, 20-33-week-old male and |
| | | female fetuses who died from spontaneous abortion; from pooled thymus tissue removed from 9, 18-32-year-old male and female donors who died from sudden death; and from pooled fetal liver tissue removed from 32, 18-24-week-old male and female fetuses. For donor A, serologies were negative. Family history included diabetes in the mother. For donor B, Serologies were negative. For donor C, pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, forming a bossellated mass replacing the right lower lobe and a portion of the middle lobe. Multiple nodules comprising the tumor show near total necrosis. Smooth muscle actin was positive. Estrogen receptor was negative and progesterone receptor was positive. The patient presented with shortness of breath. Patient history included peptic ulcer disease, normal delivery, anemia, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, hemorthoidectomy, endoscopic excision of lung lesion, and |

Table 6

| Library | Vector | Library Description |
|-----------|---------|--|
| | | <p>appendectomy. Patient medications included Megace, tamoxifen, and Pepcid. Family history included multiple sclerosis in the mother; atherosclerotic coronary artery disease and type II diabetes in the father; and breast cancer in the grandparent(s). For donor D, pathology indicated neuroectodermal tumor with advanced ganglionic differentiation. The lesion was only moderately cellular but was mitotically active with a high MIB-1 labelling index. Neuronal differentiation was widespread and advanced. Multinucleate and dysplastic-appearing forms were readily seen. The glial element was less prominent. The patient presented with motor seizures. Family history included hypertension in the grandparent(s). For donor E, pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included</p> |
| | | <p>destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father. For donor F and G, Serologies were negative for both donors and family history for donor G included diabetes in the mother.</p> |
| SKINDIA01 | PSPORT1 | This amplified library was constructed using RNA isolated from diseased skin tissue removed from 1 female and 4 males during skin biopsies. Pathologies indicated tuberculoid and lepromatous leprosy. |
| SYNWDIT01 | pINCY | Library was constructed using RNA isolated from diseased synovium tissue removed from the dorsal wrist of a 64-year-old Caucasian female. Patient history included rheumatoid arthritis. |

Table 7

| Program | Description | Reference | Parameter Threshold |
|-------------------|---|--|---|
| ABI FACTURA | A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| ABI/PARACEL FDF | A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences. | Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA. | Mismatch <50% |
| ABI AutoAssembler | A program that assembles nucleic acid sequences. | Applied Biosystems, Foster City, CA. | |
| BLAST | A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx. | Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402. | ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less |
| FASTA | A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch. | Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489. | ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater |
| BLIMPS | A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions. | Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424. | Probability value = 1.0E-3 or less |

Table 7

| Program | Description | Reference | Parameter Threshold |
|-------------|---|--|---|
| HMMER | An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM. | Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350. | PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater |
| ProfileScan | An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite. | Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221. | Normalized quality score \geq GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1. |
| Phred | A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability. | Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194. | |
| Phrap | A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences. | Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA. | Score = 120 or greater; Match length = 56 or greater |
| Consed | A graphical tool for viewing and editing Phrap assemblies. | Gordon, D. et al. (1998) Genome Res. 8:195-202. | |
| SPSscan | A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides. | Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439. | Score = 3.5 or greater |
| TMAP | A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation. | Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371. | |

Table 7

| Program | Description | Reference | Parameter Threshold |
|---------|---|---|---------------------|
| TMHMMER | A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation. | Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182. | |
| Motifs | A program that searches amino acid sequences for patterns that matched those defined in Prosite. | Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI. | |

Table 8

| SEQ ID NO: | PID | EST ID | SNP ID | EST SNP | CB1 SNP | EST Allele | Allele 1 | Allele 2 | Amino Acid | Caucasian Allele 1 frequency | African Allele 1 frequency | Asian Allele 1 frequency | Hispanic Allele 1 frequency |
|------------|---------|-----------|-------------|---------|---------|------------|----------|----------|------------|------------------------------|----------------------------|--------------------------|-----------------------------|
| 54 | 8057640 | 1319714H1 | SNP00002634 | 47 | 5294 | C | T | C | P1762 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 6862641H1 | SNP00020808 | 53 | 3738 | G | G | A | P1243 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 1991412H1 | SNP00020809 | 11 | 4035 | C | C | T | D1342 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 3626724H1 | SNP00020811 | 134 | 4308 | G | G | A | S1433 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 6328259H1 | SNP00020812 | 178 | 4320 | G | C | G | P1437 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 549375H1 | SNP00028736 | 6 | 4017 | C | C | T | D1336 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 850278H1 | SNP00028737 | 4 | 5199 | C | C | T | S1730 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 2631313H1 | SNP00028738 | 13 | 5452 | C | C | T | noncoding | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 7657674H1 | SNP00105082 | 267 | 1064 | T | C | T | L352 | n/d | n/d | n/d | n/d |
| 54 | 8057640 | 6585996H1 | SNP00115717 | 125 | 108 | C | C | A | A33 | n/a | n/a | n/a | n/a |
| 54 | 8057640 | 7164343H1 | SNP00115996 | 119 | 711 | C | C | T | G234 | 0.98 | n/d | n/d | n/d |
| 54 | 8057640 | 6603781H1 | SNP00115997 | 45 | 2861 | G | G | C | R951 | n/d | n/d | n/d | n/d |
| 54 | 8057640 | 2631313H1 | SNP00115998 | 33 | 5472 | C | C | T | noncoding | n/a | n/a | n/a | n/a |
| 55 | 7505913 | 7330339H1 | SNP00037419 | 36 | 1460 | A | A | G | noncoding | n/a | n/a | n/a | n/a |
| 55 | 7505913 | 7747030H1 | SNP00053678 | 49 | 594 | G | G | C | E117 | n/a | n/a | n/a | n/a |
| 69 | 6039650 | 4677090H1 | SNP00023512 | 16 | 2289 | C | C | T | A762 | 0.77 | 0.76 | 0.95 | 0.68 |
| 69 | 6039650 | 2790256H2 | SNP00037177 | 57 | 3774 | G | A | G | noncoding | n/a | n/a | n/a | n/a |
| 69 | 6039650 | 5959040H1 | SNP00053945 | 158 | 333 | G | A | G | G110 | n/d | n/d | n/d | n/d |
| 69 | 6039650 | 2278707H1 | SNP00053946 | 124 | 1853 | C | C | T | L617 | n/a | n/a | n/a | n/a |
| 69 | 6039650 | 5679536H1 | SNP00073169 | 76 | 2662 | T | T | C | C886 | n/a | n/a | n/a | n/a |
| 69 | 6039650 | 625807H1 | SNP00073170 | 49 | 3053 | T | T | C | C1017 | n/d | n/d | n/d | n/d |
| 69 | 6039650 | 7704080H1 | SNP00108004 | 105 | 3477 | A | A | G | E1158 | n/a | n/a | n/a | n/a |
| 69 | 6039650 | 7704080H1 | SNP00108005 | 140 | 3512 | T | T | C | W1170 | n/d | n/d | n/d | n/d |
| 69 | 6039650 | 5679536H1 | SNP00114126 | 58 | 2644 | C | C | T | G880 | n/a | n/a | n/a | n/a |
| 69 | 6039650 | 6872864H1 | SNP00114127 | 67 | 2830 | G | G | A | M942 | n/d | n/d | n/d | n/d |
| 70 | 7509919 | 7168211H1 | SNP00003102 | 65 | 381 | G | G | T | C100 | 0.71 | 0.70 | n/d | 0.76 |
| 70 | 7509919 | 7719028J1 | SNP00003103 | 36 | 1173 | G | G | A | noncoding | 0.69 | 0.80 | 0.80 | 0.57 |
| 70 | 7509919 | 3800154H1 | SNP00003104 | 15 | 1455 | A | C | A | noncoding | n/a | n/a | n/a | n/a |

Table 8

| SEQ ID NO: | PID | EST ID | SNP ID | EST SNP | CB1 SNP | EST Allele | Allele 1 | Allele 2 | Amino Acid | Caucasian Allele 1 frequency | African Allele 1 frequency | Asian Allele 1 frequency | Hispanic Allele 1 frequency |
|------------|---------|-----------|-------------|---------|---------|------------|----------|----------|------------|------------------------------|----------------------------|--------------------------|-----------------------------|
| 70 | 7509919 | 5713971H1 | SNP00003105 | 13 | 1675 | C | C | T | noncoding | n/a | n/a | n/a | n/a |
| 70 | 7509919 | 5988852H1 | SNP00003106 | 22 | 2984 | G | G | T | noncoding | n/d | n/a | n/a | n/a |
| 70 | 7509919 | 7714724J1 | SNP00021493 | 4 | 850 | A | A | G | noncoding | 0.35 | 0.63 | 0.47 | 0.37 |
| 70 | 7509919 | 4327253H1 | SNP00048150 | 115 | 2836 | G | G | C | noncoding | n/a | n/a | n/a | n/a |
| 70 | 7509919 | 7169491H1 | SNP00064963 | 42 | 110 | A | G | A | T10 | n/d | n/d | n/d | n/d |
| 71 | 7510758 | 7097756H1 | SNP00064604 | 59 | 563 | G | G | A | V143 | 0.80 | 0.94 | 0.84 | 0.86 |
| 71 | 7510758 | 7201144H2 | SNP00064605 | 14 | 623 | G | G | A | P163 | n/a | n/a | n/a | n/a |
| 71 | 7510758 | 6356584H1 | SNP00144079 | 65 | 943 | T | T | C | L270 | n/a | n/a | n/a | n/a |
| 72 | 7510063 | 6623967H1 | SNP00021188 | 54 | 1208 | C | C | T | P335 | 0.97 | n/d | n/d | n/d |
| 72 | 7510063 | 3787986H1 | SNP00021189 | 58 | 2471 | A | G | A | I756 | n/a | n/a | n/a | n/a |
| 72 | 7510063 | 6934167H1 | SNP00062632 | 41 | 1853 | T | G | T | V550 | n/a | n/a | n/a | n/a |
| 72 | 7510063 | 1594771H1 | SNP00108617 | 38 | 2202 | A | A | G | S667 | n/d | n/a | n/a | n/a |
| 72 | 7510063 | 6882193H1 | SNP00122970 | 112 | 235 | G | T | G | W11 | n/d | n/a | n/a | n/a |
| 72 | 7510063 | 6882193H1 | SNP00122971 | 232 | 355 | A | A | G | D51 | n/d | n/d | n/d | n/d |
| 72 | 7510063 | 3759933H1 | SNP00122972 | 77 | 1821 | A | A | G | I540 | n/d | n/d | n/d | n/a |
| 73 | 7510135 | 3399079H1 | SNP00001423 | 10 | 4604 | T | C | T | noncoding | 0.60 | 0.75 | 0.80 | 0.59 |
| 73 | 7510135 | 1530639H1 | SNP00017276 | 17 | 5176 | T | T | C | noncoding | n/a | n/a | n/a | n/a |
| 73 | 7510135 | 817221H1 | SNP00017277 | 23 | 5642 | C | C | T | noncoding | n/a | n/a | n/a | n/a |
| 73 | 7510135 | 4290125H1 | SNP00069160 | 10 | 4186 | T | T | C | noncoding | n/a | n/a | n/a | n/a |
| 73 | 7510135 | 6465137H1 | SNP00069161 | 26 | 4459 | G | C | G | noncoding | 0.26 | 0.21 | 0.08 | 0.26 |
| 73 | 7510135 | 7262495H1 | SNP00116903 | 37 | 1734 | T | C | T | noncoding | n/d | n/d | n/d | n/d |
| 73 | 7510135 | 6085428H1 | SNP00116904 | 154 | 1956 | A | C | A | noncoding | n/a | n/a | n/a | n/a |
| 73 | 7510135 | 3686263H1 | SNP00116905 | 12 | 4319 | C | C | A | noncoding | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 6326804H1 | SNP00021486 | 53 | 5101 | A | G | A | R1658 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 7951932H2 | SNP00056540 | 68 | 4704 | G | G | A | T1525 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 1453633H1 | SNP00061745 | 77 | 1758 | A | G | A | V543 | 0.41 | 0.47 | 0.26 | 0.49 |
| 74 | 7505011 | 4051449H1 | SNP00071405 | 10 | 5979 | C | C | A | N1950 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 6298026H1 | SNP00124562 | 88 | 4009 | A | A | G | N1294 | n/a | n/a | n/a | n/a |

Table 8

| SEQ ID NO: | PID | EST ID | SNP ID | EST SNP | CB1 SNP | EST Allele | Allele 1 | Allele 2 | Amino Acid | Caucasian Allele 1 frequency | African Allele 1 frequency | Asian Allele 1 frequency | Hispanic Allele 1 frequency |
|------------|---------|-----------|-------------|---------|---------|------------|----------|----------|------------|------------------------------|----------------------------|--------------------------|-----------------------------|
| 74 | 7505011 | 6298026H1 | SNP00124563 | 163 | 4084 | A | A | G | S1319 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 8009112H1 | SNP00130218 | 499 | 5146 | G | A | G | A1673 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 8009112H1 | SNP00130219 | 550 | 5197 | G | A | G | A1690 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 1352889H1 | SNP00130515 | 99 | 4551 | T | T | C | D1474 | n/a | n/a | n/a | n/a |
| 74 | 7505011 | 4704817H1 | SNP00130516 | 49 | 4773 | T | T | C | S1548 | n/a | n/a | n/a | n/a |

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35 and SEQ ID NO:37,
- 5 b) a polypeptide consisting essentially of an amino acid sequence of SEQ ID NO:36,
- c) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:6-7, SEQ ID NO:12-14, SEQ ID NO:19, SEQ ID NO:21, SEQ ID NO:24, SEQ ID NO:27, and SEQ ID NO:33-34,
- 10 d) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to the amino acid sequence of SEQ ID NO:9,
- e) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to the amino acid sequence of SEQ ID NO:20,
- 15 f) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:22-23,
- g) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:25,
- 20 h) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to the amino acid sequence of SEQ ID NO:32,
- i) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:16, and SEQ ID NO:35,
- 25 j) a polypeptide consisting essentially of a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-2, SEQ ID NO:4-5, SEQ ID NO:8, SEQ ID NO:10-11, SEQ ID NO:18, SEQ ID NO:26, SEQ ID NO:28-31, and SEQ ID NO:37,
- k) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, and
- 30 l) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-37.

2. An isolated polypeptide of claim 1 selected from the group consisting of:

- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35 and SEQ ID NO:37, and
- b) a polypeptide consisting essentially of the amino acid sequence of SEQ ID NO:36.

5

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

10

5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74.

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

15

7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

20

9. A method of producing a polypeptide of claim 1, the method comprising:

- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

25

10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-37.

30

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-74,

- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:38-53, SEQ ID NO:56, SEQ ID NO:60-64, SEQ ID NO:66, and SEQ ID NO:70-73,
- 5 c) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 92% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:58-59,
- d) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 93% identical to a polynucleotide sequence selected from the group consisting of
- 10 SEQ ID NO:55 and SEQ ID NO:57,
- e) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 94% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:67-68,
- f) a polynucleotide comprising a naturally occurring polynucleotide sequence at least
- 15 98% identical to the polynucleotide sequence of SEQ ID NO:69,
- g) a polynucleotide consisting essentially of a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:65 and SEQ ID NO:74,
- h) a polynucleotide complementary to a polynucleotide of a),
- 20 i) a polynucleotide complementary to a polynucleotide of b),
- j) a polynucleotide complementary to a polynucleotide of c),
- k) a polynucleotide complementary to a polynucleotide of d),
- l) a polynucleotide complementary to a polynucleotide of e),
- m) a polynucleotide complementary to a polynucleotide of f),
- 25 n) a polynucleotide complementary to a polynucleotide of g), and
- o) an RNA equivalent of a)-n).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

30

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample,

and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and

- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:
- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide is selected from the group consisting of:
- a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-35 and SEQ ID NO:37, and
- b) a polypeptide consisting essentially of the amino acid sequence of SEQ ID NO:36.

19. A method for treating a disease or condition associated with decreased expression of functional CADECM, comprising administering to a patient in need of such treatment the composition of claim 17.

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

21. A composition comprising an agonist compound identified by a method of claim 20 and a

pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional CADECM, comprising administering to a patient in need of such treatment a composition
5 of claim 21.

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 10 b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

15 25. A method for treating a disease or condition associated with overexpression of functional CADECM, comprising administering to a patient in need of such treatment a composition of claim 24.

20 26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

25

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- 30 b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the

presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,
- c) quantifying the amount of hybridization complex, and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

30. A method for a diagnostic test for a condition or disease associated with the expression of CADECM in a biological sample, the method comprising:

- a) combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- a) a chimeric antibody,

- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- e) a humanized antibody.

5

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of CADECM in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

10

34. A composition of claim 32, further comprising a label.

35. A method of diagnosing a condition or disease associated with the expression of CADECM in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

15

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

20

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from the animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37.

25

37. A polyclonal antibody produced by a method of claim 36.

30

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide consisting of an amino acid sequence selected from the group consisting of SEQ ID NO:1-37, or an immunogenic fragment

35

thereof, under conditions to elicit an antibody response,

- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- 5 d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37.

10 40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

15 42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

20 44. A method of detecting a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- 25 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37 in the sample.

45. A method of purifying a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37 from a sample, the method comprising:

- 30 a) incubating the antibody of claim 11 with the sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-37.

35

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating an expression profile of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains

multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

5

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

10

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

15

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

20

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

64. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.

65. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.

25

66. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.

67. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.

30

68. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.

69. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.

70. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.

35

71. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:16.
72. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:17.
- 5 73. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:18.
74. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:19.
75. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:20.
- 10 76. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:21.
77. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:22.
- 15 78. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:23.
79. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:24.
80. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:25.
- 20 81. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:26.
82. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:27.
- 25 83. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:28.
84. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:29.
85. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:30.
- 30 86. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:31.
87. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:32.
- 35 88. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:33.

89. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:34.
90. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:35.
- 5 91. A polypeptide of claim 1, consisting essentially of the amino acid sequence of SEQ ID NO:36.
92. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:37.
- 10 93. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:38.
94. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:39.
- 15 95. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:40.
96. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:41.
- 20 97. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:42.
98. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:43.
- 25 99. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:44.
- 30 100. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:45.
101. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:46.
- 35

102. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:47.

103. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
5 NO:48.

104. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:49.

105. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
10 NO:50.

106. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:51.

15 107. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:52.

108. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
20 NO:53.

109. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:54.

25 110. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:55.

111. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:56.

30 112. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
NO:57.

113. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID
35 NO:58.

114. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:59.

115. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:60.

116. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:61.

117. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:62.

118. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:63.

119. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:64.

120. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:65.

121. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:66.

122. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:67.

123. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:68.

124. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:69.

125. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:70.

126. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:71.

5 127. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:72.

128. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:73.

10 129. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:74.

<110> INCYTE GENOMICS, INC.
BAUGHN, Mariah R.
BECHA, Shanya D.
BHATIA, Umesh G.
BLAKE, Julie J.
BOROWSKY, Mark L.
BURRILL, John D.
DELEGEANE, Angelo M.
ELLIOTT, Vicki S.
GANDHI, Ameena R.
GIETZEN, Kimberly J.
GORVAD, Ann E.
GRIFFIN, Jennifer A.
HO, Anne
JIN, Pei
KABLE, Amy E.
LAL, Preeti G.
LEE, Ernestine A.
LEE, Sally
LEE, Soo Yuen
MARQUIS, Joseph P.
LEHR-MASON, Patricia M.
RAMKUMAR, Jayalaxmi
RICHARDSON, Thomas W.
SPRAGUE, William W.
SWARNAKAR, Anita
TANG, Y. Tom
TRAN, Bao
TRAN, Uyen K.
CHAWLA, Narinder K.
WARREN, Bridget A.
XU, Yuming
YUE, Henry
ZHENG, Wenjin

<120> CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

<130> PF-1296 PCT

<140> To Be Assigned

<141> Herewith

<150> US 60/334,343

<151> 2001-11-30

<150> US 60/340,278

<151> 2001-12-07

<150> US 60/345,069

<151> 2002-01-04

<150> US 60/351,352

<151> 2002-01-25

<150> US 60/357,168

<151> 2002-02-14

<150> US 60/369,128

<151> 2002-03-29

<150> US 60/370,802

<151> 2002-04-05

<160> 74
 <170> PERL Program

<210> 1
 <211> 311
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7504868CD1

<400> 1
 Met Asp Phe Gly Leu Ala Leu Leu Leu Ala Gly Leu Leu Gly Leu
 1 5 10 15
 Leu Leu Ala Phe Pro Asp Gln Leu Thr Val Ser Pro Ala Ala Leu
 20 25 30
 Val Pro Gly Asp Pro Glu Val Ala Cys Thr Ala His Lys Val Thr
 35 40 45
 Pro Val Asp Pro Asn Ala Leu Ser Phe Ser Leu Leu Val Gly Gly
 50 55 60
 Gln Glu Leu Glu Gly Ala Gln Ala Leu Gly Pro Glu Val Gln Glu
 65 70 75
 Glu Glu Glu Glu Pro Gln Gly Asp Glu Asp Val Leu Phe Arg Val
 80 85 90
 Thr Glu Arg Trp Arg Leu Pro Pro Leu Gly Thr Pro Val Pro Pro
 95 100 105
 Ala Leu Tyr Cys Gln Ala Thr Met Arg Leu Pro Gly Leu Glu Leu
 110 115 120
 Ser His Arg Gln Ala Ile Pro Val Leu His Ser Pro Thr Ser Pro
 125 130 135
 Glu Pro Pro Asp Thr Thr Ser Pro Glu Pro Pro Asn Thr Thr Ser
 140 145 150
 Pro Glu Ser Pro Asp Thr Thr Ser Pro Glu Ser Pro Asp Thr Thr
 155 160 165
 Ser Gln Glu Pro Pro Asp Thr Thr Ser Gln Glu Pro Pro Asp Thr
 170 175 180
 Thr Ser Gln Glu Pro Pro Asp Thr Thr Ser Pro Glu Pro Pro Asp
 185 190 195
 Lys Thr Ser Pro Glu Pro Ala Pro Gln Gln Gly Ser Thr His Thr
 200 205 210
 Pro Arg Ser Pro Gly Ser Thr Arg Thr Arg Arg Pro Glu Ile Ser
 215 220 225
 Gln Ala Gly Pro Thr Gln Gly Glu Val Ile Pro Thr Gly Ser Ser
 230 235 240
 Lys Pro Ala Gly Asp Gln Leu Pro Ala Ala Leu Trp Thr Ser Ser
 245 250 255
 Ala Val Leu Gly Leu Leu Leu Leu Ala Leu Pro Thr Tyr His Leu
 260 265 270
 Trp Lys Arg Cys Arg His Leu Ala Glu Asp Asp Thr His Pro Pro
 275 280 285
 Ala Ser Leu Arg Leu Leu Pro Gln Val Ser Ala Trp Ala Gly Leu
 290 295 300
 Arg Gly Thr Gly Gln Val Gly Ile Ser Pro Ser
 305 310

<210> 2
 <211> 1445
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature

<223> Incyte ID No: 7504930CD1

<400> 2

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Met | Ala | Ala | Glu | Arg | Gly | Ala | Arg | Arg | Leu | Leu | Ser | Thr | Pro | Ser | |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Phe | Trp | Leu | Tyr | Cys | Leu | Leu | Leu | Leu | Gly | Arg | Arg | Ala | Pro | Gly | |
| | | | | 20 | | | | | 25 | | | | | 30 | |
| Ala | Ala | Ala | Ala | Arg | Ser | Gly | Ser | Ala | Pro | Gln | Ser | Pro | Gly | Ala | |
| | | | | 35 | | | | | 40 | | | | | 45 | |
| Ser | Ile | Arg | Thr | Phe | Thr | Pro | Phe | Tyr | Phe | Leu | Val | Glu | Pro | Val | |
| | | | | 50 | | | | | 55 | | | | | 60 | |
| Asp | Thr | Leu | Ser | Val | Arg | Gly | Ser | Ser | Val | Ile | Leu | Asn | Cys | Ser | |
| | | | | 65 | | | | | 70 | | | | | 75 | |
| Ala | Tyr | Ser | Glu | Pro | Ser | Pro | Lys | Ile | Glu | Trp | Lys | Lys | Asp | Gly | |
| | | | | 80 | | | | | 85 | | | | | 90 | |
| Thr | Phe | Leu | Asn | Leu | Val | Ser | Asp | Asp | Arg | Arg | Gln | Leu | Leu | Pro | |
| | | | | 95 | | | | | 100 | | | | | 105 | |
| Asp | Gly | Ser | Leu | Phe | Ile | Ser | Asn | Val | Val | His | Ser | Lys | His | Asn | |
| | | | | 110 | | | | | 115 | | | | | 120 | |
| Lys | Pro | Asp | Glu | Gly | Tyr | Tyr | Gln | Cys | Val | Ala | Thr | Val | Glu | Ser | |
| | | | | 125 | | | | | 130 | | | | | 135 | |
| Leu | Gly | Thr | Ile | Ile | Ser | Arg | Thr | Ala | Lys | Leu | Ile | Val | Ala | Gly | |
| | | | | 140 | | | | | 145 | | | | | 150 | |
| Leu | Pro | Arg | Phe | Thr | Ser | Gln | Pro | Glu | Pro | Ser | Ser | Val | Tyr | Ala | |
| | | | | 155 | | | | | 160 | | | | | 165 | |
| Gly | Asn | Asn | Ala | Ile | Leu | Asn | Cys | Glu | Val | Asn | Ala | Asp | Leu | Val | |
| | | | | 170 | | | | | 175 | | | | | 180 | |
| Pro | Phe | Val | Arg | Trp | Glu | Gln | Asn | Arg | Gln | Pro | Leu | Leu | Leu | Asp | |
| | | | | 185 | | | | | 190 | | | | | 195 | |
| Asp | Arg | Val | Ile | Lys | Leu | Pro | Ser | Gly | Met | Leu | Val | Ile | Ser | Asn | |
| | | | | 200 | | | | | 205 | | | | | 210 | |
| Ala | Thr | Glu | Gly | Asp | Gly | Gly | Leu | Tyr | Arg | Cys | Val | Val | Glu | Ser | |
| | | | | 215 | | | | | 220 | | | | | 225 | |
| Gly | Gly | Pro | Pro | Lys | Tyr | Ser | Asp | Glu | Val | Glu | Leu | Lys | Val | Leu | |
| | | | | 230 | | | | | 235 | | | | | 240 | |
| Pro | Asp | Pro | Glu | Val | Ile | Ser | Asp | Leu | Val | Phe | Leu | Lys | Gln | Pro | |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Ser | Pro | Leu | Val | Arg | Val | Ile | Gly | Gln | Asp | Val | Val | Leu | Pro | Cys | |
| | | | | 260 | | | | | 265 | | | | | 270 | |
| Val | Ala | Ser | Gly | Leu | Pro | Thr | Pro | Thr | Ile | Lys | Trp | Met | Lys | Asn | |
| | | | | 275 | | | | | 280 | | | | | 285 | |
| Glu | Glu | Ala | Leu | Asp | Thr | Glu | Ser | Ser | Glu | Arg | Leu | Val | Leu | Leu | |
| | | | | 290 | | | | | 295 | | | | | 300 | |
| Ala | Gly | Gly | Ser | Leu | Glu | Ile | Ser | Asp | Val | Thr | Glu | Asp | Asp | Ala | |
| | | | | 305 | | | | | 310 | | | | | 315 | |
| Gly | Thr | Tyr | Phe | Cys | Ile | Ala | Asp | Asn | Gly | Asn | Glu | Thr | Ile | Glu | |
| | | | | 320 | | | | | 325 | | | | | 330 | |
| Ala | Gln | Ala | Glu | Leu | Thr | Val | Gln | Ala | Gln | Pro | Glu | Phe | Leu | Lys | |
| | | | | 335 | | | | | 340 | | | | | 345 | |
| Gln | Pro | Thr | Asn | Ile | Tyr | Ala | His | Glu | Ser | Met | Asp | Ile | Val | Phe | |
| | | | | 350 | | | | | 355 | | | | | 360 | |
| Glu | Cys | Glu | Val | Thr | Gly | Lys | Pro | Thr | Pro | Thr | Val | Lys | Trp | Val | |
| | | | | 365 | | | | | 370 | | | | | 375 | |
| Lys | Asn | Gly | Asp | Met | Val | Ile | Pro | Ser | Asp | Tyr | Phe | Lys | Ile | Val | |
| | | | | 380 | | | | | 385 | | | | | 390 | |
| Lys | Glu | His | Asn | Leu | Gln | Val | Leu | Gly | Leu | Val | Lys | Ser | Asp | Glu | |
| | | | | 395 | | | | | 400 | | | | | 405 | |
| Gly | Phe | Tyr | Gln | Cys | Ile | Ala | Glu | Asn | Asp | Val | Gly | Asn | Ala | Gln | |
| | | | | 410 | | | | | 415 | | | | | 420 | |
| Ala | Gly | Ala | Gln | Leu | Ile | Ile | Leu | Glu | His | Ala | Pro | Ala | Thr | Thr | |
| | | | | 425 | | | | | 430 | | | | | 435 | |
| Gly | Pro | Leu | Pro | Ser | Ala | Pro | Arg | Asp | Val | Val | Ala | Ser | Leu | Val | |
| | | | | 440 | | | | | 445 | | | | | 450 | |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ser | Thr | Arg | Phe | Ile | Lys | Leu | Thr | Trp | Arg | Thr | Pro | Ala | Ser | Asp |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Pro | His | Gly | Asp | Asn | Leu | Thr | Tyr | Ser | Val | Phe | Tyr | Thr | Lys | Glu |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Gly | Ile | Ala | Arg | Glu | Arg | Val | Glu | Asn | Thr | Ser | His | Pro | Gly | Glu |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Met | Gln | Val | Thr | Ile | Gln | Asn | Leu | Met | Pro | Ala | Thr | Val | Tyr | Ile |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Phe | Arg | Val | Met | Ala | Gln | Asn | Lys | His | Gly | Ser | Gly | Glu | Ser | Ser |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Ala | Pro | Leu | Arg | Val | Glu | Thr | Gln | Pro | Glu | Val | Gln | Leu | Pro | Gly |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Pro | Ala | Pro | Asn | Leu | Arg | Ala | Tyr | Ala | Ala | Ser | Pro | Thr | Ser | Ile |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Thr | Val | Thr | Trp | Glu | Thr | Pro | Val | Ser | Gly | Asn | Gly | Glu | Ile | Gln |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Asn | Tyr | Lys | Leu | Tyr | Tyr | Met | Glu | Lys | Gly | Thr | Asp | Lys | Glu | Gln |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Asp | Val | Asp | Val | Ser | Ser | His | Ser | Tyr | Thr | Ile | Asn | Gly | Leu | Lys |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Lys | Tyr | Thr | Glu | Tyr | Ser | Phe | Arg | Val | Val | Ala | Tyr | Asn | Lys | His |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Gly | Pro | Gly | Val | Ser | Thr | Pro | Asp | Val | Ala | Val | Arg | Thr | Leu | Ser |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Asp | Val | Pro | Ser | Ala | Ala | Pro | Gln | Asn | Leu | Ser | Leu | Glu | Val | Arg |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Asn | Ser | Lys | Ser | Ile | Met | Ile | His | Trp | Gln | Pro | Pro | Ala | Pro | Ala |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Thr | Gln | Asn | Gly | Gln | Ile | Thr | Gly | Tyr | Lys | Ile | Arg | Tyr | Arg | Lys |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Ala | Ser | Arg | Lys | Ser | Asp | Val | Thr | Glu | Thr | Leu | Val | Ser | Gly | Thr |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Gln | Leu | Ser | Gln | Leu | Ile | Glu | Gly | Leu | Asp | Arg | Gly | Thr | Glu | Tyr |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Asn | Phe | Arg | Val | Ala | Ala | Leu | Thr | Ile | Asn | Gly | Thr | Gly | Pro | Ala |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Thr | Asp | Trp | Leu | Ser | Ala | Glu | Thr | Phe | Glu | Ser | Asp | Leu | Asp | Glu |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Thr | Arg | Val | Pro | Glu | Val | Pro | Ser | Ser | Leu | His | Val | Arg | Pro | Leu |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Val | Thr | Ser | Ile | Val | Val | Ser | Trp | Thr | Pro | Pro | Glu | Asn | Gln | Asn |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Ile | Val | Val | Arg | Gly | Tyr | Ala | Ile | Gly | Tyr | Gly | Ile | Gly | Ser | Pro |
| | | | | 770 | | | | | 775 | | | | | 780 |
| His | Ala | Gln | Thr | Ile | Lys | Val | Asp | Tyr | Lys | Gln | Arg | Tyr | Tyr | Thr |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Ile | Glu | Asn | Leu | Asp | Pro | Ser | Ser | His | Tyr | Val | Ile | Thr | Leu | Lys |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Ala | Phe | Asn | Asn | Val | Gly | Glu | Gly | Ile | Pro | Leu | Tyr | Glu | Ser | Ala |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Val | Thr | Arg | Pro | His | Thr | Val | Pro | Asp | Pro | Thr | Pro | Met | Met | Pro |
| | | | | 830 | | | | | 835 | | | | | 840 |
| Pro | Val | Gly | Val | Gln | Ala | Ser | Ile | Leu | Ser | His | Asp | Thr | Ile | Arg |
| | | | | 845 | | | | | 850 | | | | | 855 |
| Ile | Thr | Trp | Ala | Asp | Asn | Ser | Leu | Pro | Lys | His | Gln | Lys | Ile | Thr |
| | | | | 860 | | | | | 865 | | | | | 870 |
| Asp | Ser | Arg | Tyr | Tyr | Thr | Val | Arg | Trp | Lys | Thr | Asn | Ile | Pro | Ala |
| | | | | 875 | | | | | 880 | | | | | 885 |
| Asn | Thr | Lys | Tyr | Lys | Asn | Ala | Asn | Ala | Thr | Thr | Leu | Ser | Tyr | Leu |
| | | | | 890 | | | | | 895 | | | | | 900 |
| Val | Thr | Gly | Leu | Lys | Pro | Asn | Thr | Leu | Tyr | Glu | Phe | Ser | Val | Met |
| | | | | 905 | | | | | 910 | | | | | 915 |
| Val | Thr | Lys | Gly | Arg | Arg | Ser | Ser | Thr | Trp | Ser | Met | Thr | Ala | His |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|
| | | | | 920 | | | | | 925 | | | | 930 | |
| Gly | Thr | Thr | Phe | Glu | Leu | Val | Pro | Thr | Ser | Pro | Pro | Lys | Asp | Val |
| | | | | 935 | | | | | 940 | | | | | 945 |
| Thr | Val | Val | Ser | Lys | Glu | Gly | Lys | Pro | Lys | Thr | Ile | Ile | Val | Asn |
| | | | | 950 | | | | | 955 | | | | | 960 |
| Trp | Gln | Pro | Pro | Ser | Glu | Ala | Asn | Gly | Lys | Ile | Thr | Gly | Tyr | Ile |
| | | | | 965 | | | | | 970 | | | | | 975 |
| Ile | Tyr | Tyr | Ser | Thr | Asp | Val | Asn | Ala | Glu | Ile | His | Asp | Trp | Val |
| | | | | 980 | | | | | 985 | | | | | 990 |
| Ile | Glu | Pro | Val | Val | Gly | Asn | Arg | Leu | Thr | His | Gln | Ile | Gln | Glu |
| | | | | 995 | | | | | 1000 | | | | | 1005 |
| Leu | Thr | Leu | Asp | Thr | Pro | Tyr | Tyr | Phe | Lys | Ile | Gln | Ala | Arg | Asn |
| | | | | 1010 | | | | | 1015 | | | | | 1020 |
| Ser | Lys | Gly | Met | Gly | Pro | Met | Ser | Glu | Ala | Val | Gln | Phe | Arg | Thr |
| | | | | 1025 | | | | | 1030 | | | | | 1035 |
| Pro | Lys | Ala | Asp | Ser | Ser | Asp | Lys | Met | Pro | Asn | Asp | Gln | Ala | Ser |
| | | | | 1040 | | | | | 1045 | | | | | 1050 |
| Gly | Ser | Gly | Gly | Lys | Gly | Ser | Arg | Leu | Pro | Asp | Leu | Gly | Ser | Asp |
| | | | | 1055 | | | | | 1060 | | | | | 1065 |
| Tyr | Lys | Pro | Pro | Met | Ser | Gly | Ser | Asn | Ser | Pro | His | Gly | Ser | Pro |
| | | | | 1070 | | | | | 1075 | | | | | 1080 |
| Thr | Ser | Pro | Leu | Asp | Ser | Asn | Met | Leu | Leu | Val | Ile | Ile | Val | Ser |
| | | | | 1085 | | | | | 1090 | | | | | 1095 |
| Val | Gly | Val | Ile | Thr | Ile | Val | Val | Val | Val | Ile | Ile | Ala | Val | Phe |
| | | | | 1100 | | | | | 1105 | | | | | 1110 |
| Cys | Thr | Arg | Arg | Thr | Thr | Ser | His | Gln | Lys | Lys | Lys | Arg | Ala | Ala |
| | | | | 1115 | | | | | 1120 | | | | | 1125 |
| Cys | Lys | Ser | Val | Asn | Gly | Ser | His | Lys | Tyr | Lys | Gly | Asn | Ser | Lys |
| | | | | 1130 | | | | | 1135 | | | | | 1140 |
| Asp | Val | Lys | Pro | Pro | Asp | Leu | Trp | Ile | His | His | Glu | Arg | Leu | Glu |
| | | | | 1145 | | | | | 1150 | | | | | 1155 |
| Leu | Lys | Pro | Ile | Asp | Lys | Ser | Pro | Asp | Pro | Asn | Pro | Ile | Met | Thr |
| | | | | 1160 | | | | | 1165 | | | | | 1170 |
| Asp | Thr | Pro | Ile | Pro | Arg | Asn | Ser | Gln | Asp | Ile | Thr | Pro | Val | Asp |
| | | | | 1175 | | | | | 1180 | | | | | 1185 |
| Asn | Ser | Met | Asp | Ser | Asn | Ile | His | Gln | Arg | Arg | Asn | Ser | Tyr | Arg |
| | | | | 1190 | | | | | 1195 | | | | | 1200 |
| Gly | His | Glu | Ser | Glu | Asp | Ser | Met | Ser | Thr | Leu | Ala | Gly | Arg | Arg |
| | | | | 1205 | | | | | 1210 | | | | | 1215 |
| Gly | Met | Arg | Pro | Lys | Met | Met | Met | Pro | Phe | Asp | Ser | Gln | Pro | Pro |
| | | | | 1220 | | | | | 1225 | | | | | 1230 |
| Gln | Pro | Val | Ile | Ser | Ala | His | Pro | Ile | His | Ser | Leu | Asp | Asn | Pro |
| | | | | 1235 | | | | | 1240 | | | | | 1245 |
| His | His | His | Phe | His | Ser | Ser | Ser | Leu | Ala | Ser | Pro | Ala | Arg | Ser |
| | | | | 1250 | | | | | 1255 | | | | | 1260 |
| His | Leu | Tyr | His | Pro | Gly | Ser | Pro | Trp | Pro | Ile | Gly | Thr | Ser | Met |
| | | | | 1265 | | | | | 1270 | | | | | 1275 |
| Ser | Leu | Ser | Asp | Arg | Ala | Asn | Ser | Thr | Glu | Ser | Val | Arg | Asn | Thr |
| | | | | 1280 | | | | | 1285 | | | | | 1290 |
| Pro | Ser | Thr | Asp | Thr | Met | Pro | Ala | Ser | Ser | Ser | Gln | Thr | Cys | Cys |
| | | | | 1295 | | | | | 1300 | | | | | 1305 |
| Thr | Asp | His | Gln | Asp | Pro | Glu | Gly | Ala | Thr | Ser | Ser | Ser | Tyr | Leu |
| | | | | 1310 | | | | | 1315 | | | | | 1320 |
| Ala | Ser | Ser | Gln | Glu | Glu | Asp | Ser | Gly | Gln | Ser | Leu | Pro | Thr | Ala |
| | | | | 1325 | | | | | 1330 | | | | | 1335 |
| His | Val | Arg | Pro | Ser | His | Pro | Leu | Lys | Ser | Phe | Ala | Val | Pro | Ala |
| | | | | 1340 | | | | | 1345 | | | | | 1350 |
| Ile | Pro | Pro | Pro | Gly | Pro | Pro | Thr | Tyr | Asp | Pro | Ala | Leu | Pro | Ser |
| | | | | 1355 | | | | | 1360 | | | | | 1365 |
| Thr | Pro | Leu | Leu | Ser | Gln | Gln | Ala | Leu | Asn | His | His | Ile | His | Ser |
| | | | | 1370 | | | | | 1375 | | | | | 1380 |
| Val | Lys | Thr | Ala | Ser | Ile | Gly | Thr | Leu | Gly | Arg | Ser | Arg | Pro | Pro |
| | | | | 1385 | | | | | 1390 | | | | | 1395 |

```

Met Pro Val Val Val Pro Ser Ala Pro Glu Val Gln Glu Thr Thr
      1400      1405      1410
Arg Met Leu Glu Asp Ser Glu Ser Ser Tyr Glu Pro Asp Glu Leu
      1415      1420      1425
Thr Lys Glu Met Ala His Leu Glu Gly Leu Met Lys Asp Leu Asn
      1430      1435      1440
Ala Ile Thr Thr Ala
      1445

```

```

<210> 3
<211> 726
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 6610456CD1

```

```

<400> 3
Met Ala Gln Thr Val Gln Asn Val Thr Leu Ser Leu Thr Leu Pro
  1      5      10      15
Ile Thr Cys His Ile Cys Leu Gly Lys Val Arg Gln Pro Val Ile
      20      25      30
Cys Ile Asn Asn His Val Phe Cys Ser Ile Cys Ile Asp Leu Trp
      35      40      45
Leu Lys Asn Asn Ser Gln Cys Pro Ala Cys Arg Val Pro Ile Thr
      50      55      60
Pro Glu Asn Pro Cys Lys Glu Ile Ile Gly Thr Ser Glu Ser
      65      70      75
Glu Pro Met Leu Ser His Thr Val Arg Lys His Leu Arg Lys Thr
      80      85      90
Arg Leu Glu Leu Leu His Lys Glu Tyr Glu Asp Glu Ile Asp Cys
      95      100      105
Leu Gln Lys Glu Val Glu Glu Leu Lys Ser Lys Asn Leu Ser Leu
      110      115      120
Glu Ser Gln Ile Lys Thr Ile Leu Asp Pro Leu Thr Leu Val Gln
      125      130      135
Gly Asn Gln Asn Glu Asp Lys His Leu Val Thr Asp Asn Pro Ser
      140      145      150
Lys Ile Asn Pro Glu Thr Val Ala Glu Trp Lys Lys Lys Leu Arg
      155      160      165
Thr Ala Asn Glu Ile Tyr Glu Lys Val Lys Asp Asp Val Asp Lys
      170      175      180
Leu Lys Glu Ala Asn Lys Lys Leu Lys Leu Glu Asn Gly Gly Leu
      185      190      195
Val Arg Glu Asn Leu Arg Leu Lys Ala Glu Val Asp Asn Arg Ser
      200      205      210
Pro Gln Lys Phe Gly Arg Phe Ala Val Ala Ala Leu Gln Ser Lys
      215      220      225
Val Glu Gln Tyr Glu Arg Glu Thr Asn Arg Leu Lys Lys Ala Leu
      230      235      240
Glu Arg Ser Asp Lys Tyr Ile Glu Glu Leu Glu Ser Gln Val Ala
      245      250      255
Gln Leu Lys Asn Ser Ser Glu Glu Lys Glu Ala Met Asn Ser Ile
      260      265      270
Cys Gln Thr Ala Leu Ser Ala Asp Gly Lys Gly Ser Lys Gly Ser
      275      280      285
Glu Glu Asp Val Val Ser Lys Asn Gln Gly Asp Ser Ala Arg Lys
      290      295      300
Gln Pro Gly Ser Ser Thr Ser Ser Ser Ser His Leu Ala Lys Pro
      305      310      315
Ser Ser Ser Arg Leu Cys Asp Thr Ser Ser Ala Arg Gln Glu Ser
      320      325      330

```

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Ser | Lys | Ala | Asp | Leu | Asn | Cys | Ser | Lys | Asn | Lys | Asp | Leu | Tyr | 335 | 340 | 345 |
| Gln | Glu | Gln | Val | Glu | Val | Met | Leu | Asp | Val | Thr | Asp | Thr | Ser | Met | 350 | 355 | 360 |
| Asp | Thr | Tyr | Leu | Glu | Arg | Glu | Trp | Gly | Asn | Lys | Pro | Ser | Asp | Cys | 365 | 370 | 375 |
| Val | Pro | Tyr | Lys | Asp | Glu | Glu | Leu | Tyr | Asp | Leu | Pro | Ala | Pro | Cys | 380 | 385 | 390 |
| Thr | Pro | Leu | Ser | Leu | Ser | Cys | Leu | Gln | Leu | Ser | Thr | Pro | Glu | Asn | 395 | 400 | 405 |
| Arg | Glu | Ser | Ser | Val | Val | Gln | Ala | Gly | Gly | Ser | Lys | Lys | His | Ser | 410 | 415 | 420 |
| Asn | His | Leu | Arg | Lys | Leu | Val | Phe | Asp | Asp | Phe | Cys | Asp | Ser | Ser | 425 | 430 | 435 |
| Asn | Val | Ser | Asn | Lys | Asp | Ser | Ser | Glu | Asp | Asp | Ile | Ser | Arg | Ser | 440 | 445 | 450 |
| Glu | Asn | Glu | Lys | Lys | Ser | Glu | Cys | Phe | Ser | Ser | Pro | Lys | Thr | Gly | 455 | 460 | 465 |
| Phe | Trp | Asp | Cys | Cys | Ser | Thr | Ser | Tyr | Ala | Gln | Asn | Leu | Asp | Phe | 470 | 475 | 480 |
| Glu | Ser | Ser | Glu | Gly | Asn | Thr | Ile | Ala | Asn | Ser | Val | Gly | Glu | Ile | 485 | 490 | 495 |
| Ser | Ser | Lys | Leu | Ser | Glu | Lys | Ser | Gly | Leu | Cys | Leu | Ser | Lys | Arg | 500 | 505 | 510 |
| Leu | Asn | Ser | Ile | Arg | Ser | Phe | Glu | Met | Asn | Arg | Thr | Arg | Thr | Ser | 515 | 520 | 525 |
| Ser | Glu | Ala | Ser | Met | Asp | Ala | Ala | Tyr | Leu | Asp | Lys | Ile | Ser | Glu | 530 | 535 | 540 |
| Leu | Asp | Ser | Met | Met | Ser | Glu | Ser | Asp | Asn | Ser | Lys | Ser | Pro | Cys | 545 | 550 | 555 |
| Asn | Asn | Gly | Phe | Lys | Ser | Leu | Asp | Leu | Asp | Gly | Leu | Ser | Lys | Ser | 560 | 565 | 570 |
| Ser | Gln | Gly | Ser | Glu | Phe | Leu | Glu | Glu | Pro | Asp | Lys | Leu | Glu | Glu | 575 | 580 | 585 |
| Lys | Thr | Glu | Leu | Asn | Leu | Ser | Lys | Gly | Ser | Leu | Thr | Asn | Asp | Gln | 590 | 595 | 600 |
| Leu | Glu | Asn | Gly | Ser | Glu | Trp | Lys | Pro | Thr | Ser | Phe | Phe | Leu | Leu | 605 | 610 | 615 |
| Ser | Pro | Ser | Asp | Gln | Glu | Met | Asn | Glu | Asp | Phe | Ser | Leu | His | Ser | 620 | 625 | 630 |
| Ser | Ser | Cys | Pro | Val | Thr | Asn | Glu | Ile | Lys | Pro | Pro | Ser | Cys | Leu | 635 | 640 | 645 |
| Phe | Gln | Thr | Glu | Phe | Ser | Gln | Gly | Ile | Leu | Leu | Ser | Ser | Ser | His | 650 | 655 | 660 |
| Arg | Leu | Phe | Glu | Asp | Gln | Arg | Phe | Gly | Ser | Ser | Leu | Phe | Lys | Met | 665 | 670 | 675 |
| Ser | Ser | Glu | Met | His | Ser | Leu | His | Asn | His | Leu | Gln | Ser | Pro | Trp | 680 | 685 | 690 |
| Ser | Thr | Ser | Phe | Val | Pro | Glu | Lys | Arg | Asn | Lys | Asn | Val | Asn | Gln | 695 | 700 | 705 |
| Ser | Thr | Lys | Arg | Lys | Ile | Gln | Ser | Ser | Leu | Ser | Ser | Ala | Ser | Pro | 710 | 715 | 720 |
| Ser | Lys | Ala | Thr | Lys | Ser | | | | | | | | | | 725 | | |

<210> 4

<211> 1474

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503573CD1

<400> 4

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Gly | Asp | Val | Lys | Ala | Leu | Leu | Phe | Val | Ala | Ala | Ala | Arg | Ala |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Arg | Arg | Leu | Gly | Gly | Ala | Ala | Ala | Ser | Glu | Ser | Leu | Ala | Val | Ser |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Glu | Ala | Phe | Cys | Arg | Val | Arg | Ser | Cys | Gln | Pro | Lys | Lys | Cys | Ala |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Gly | Pro | Gln | Arg | Cys | Leu | Asn | Pro | Val | Pro | Ala | Val | Pro | Ser | Pro |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Ser | Pro | Ser | Val | Arg | Lys | Arg | Gln | Val | Ser | Leu | Asn | Trp | Gln | Pro |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Leu | Thr | Leu | Gln | Glu | Ala | Arg | Ala | Leu | Leu | Lys | Arg | Arg | Arg | Pro |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Arg | Gly | Pro | Gly | Gly | Arg | Gly | Leu | Leu | Arg | Arg | Arg | Pro | Pro | Gln |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Arg | Ala | Pro | Ala | Gly | Lys | Ala | Pro | Val | Leu | Cys | Pro | Leu | Ile | Cys |
| | | | | 110 | | | | | 115 | | | | | 120 |
| His | Asn | Gly | Gly | Val | Cys | Val | Lys | Pro | Asp | Arg | Cys | Leu | Cys | Pro |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Pro | Asp | Phe | Ala | Gly | Lys | Phe | Cys | Gln | Leu | His | Ser | Ser | Gly | Ala |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Arg | Pro | Pro | Ala | Pro | Ala | Val | Pro | Gly | Leu | Thr | Arg | Ser | Val | Tyr |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Thr | Met | Pro | Leu | Ala | Asn | His | Arg | Asp | Asp | Glu | His | Gly | Val | Ala |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Ser | Met | Val | Ser | Val | His | Val | Glu | His | Pro | Gln | Glu | Ala | Ser | Val |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Val | Val | His | Gln | Val | Glu | Arg | Val | Ser | Gly | Pro | Trp | Glu | Glu | Ala |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Asp | Ala | Glu | Ala | Val | Ala | Arg | Ala | Glu | Ala | Ala | Ala | Arg | Ala | Glu |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Ala | Ala | Ala | Pro | Tyr | Thr | Val | Leu | Ala | Gln | Ser | Ala | Pro | Arg | Glu |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Asp | Gly | Tyr | Ser | Asp | Ala | Ser | Gly | Phe | Gly | Tyr | Cys | Phe | Arg | Glu |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Leu | Arg | Gly | Gly | Glu | Cys | Ala | Ser | Pro | Leu | Pro | Gly | Leu | Arg | Thr |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Gln | Glu | Val | Cys | Cys | Arg | Gly | Ala | Gly | Leu | Ala | Trp | Gly | Val | His |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asp | Cys | Gln | Leu | Cys | Ser | Glu | Arg | Leu | Gly | Asn | Ser | Glu | Arg | Val |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Ser | Ala | Pro | Asp | Gly | Pro | Cys | Pro | Thr | Gly | Phe | Glu | Arg | Val | Asn |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Gly | Ser | Cys | Glu | Asp | Val | Asp | Glu | Cys | Ala | Thr | Gly | Gly | Arg | Cys |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Gln | His | Gly | Glu | Cys | Ala | Asn | Thr | Arg | Gly | Gly | Tyr | Thr | Cys | Val |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Cys | Pro | Asp | Gly | Phe | Leu | Leu | Asp | Ser | Ser | Arg | Ser | Ser | Cys | Ile |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Ser | Gln | His | Val | Ile | Ser | Glu | Ala | Lys | Gly | Pro | Cys | Phe | Arg | Val |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Leu | Arg | Asp | Gly | Gly | Cys | Ser | Leu | Pro | Ile | Leu | Arg | Asn | Ile | Thr |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Lys | Gln | Ile | Cys | Cys | Cys | Ser | Arg | Val | Gly | Lys | Ala | Trp | Gly | Arg |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Gly | Cys | Gln | Leu | Cys | Pro | Pro | Phe | Gly | Ser | Glu | Gly | Phe | Arg | Glu |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Ile | Cys | Pro | Ala | Gly | Pro | Gly | Tyr | His | Tyr | Ser | Ala | Ser | Asp | Leu |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Arg | Tyr | Asn | Thr | Arg | Pro | Leu | Gly | Gln | Glu | Pro | Pro | Arg | Val | Ser |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Leu | Ser | Gln | Pro | Arg | Thr | Leu | Pro | Ala | Thr | Ser | Arg | Pro | Ser | Ala |
| | | | | 455 | | | | | 460 | | | | | 465 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Phe | Leu | Pro | Thr | His | Arg | Leu | Glu | Pro | Arg | Pro | Glu | Pro | Arg |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Pro | Asp | Pro | Arg | Pro | Gly | Pro | Glu | Leu | Pro | Leu | Pro | Ser | Ile | Pro |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Ala | Trp | Thr | Gly | Pro | Glu | Ile | Pro | Glu | Ser | Gly | Pro | Ser | Ser | Gly |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Met | Cys | Gln | Arg | Asn | Pro | Gln | Val | Cys | Gly | Pro | Gly | Arg | Cys | Ile |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Ser | Arg | Pro | Ser | Gly | Tyr | Thr | Cys | Ala | Cys | Asp | Ser | Gly | Phe | Arg |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Leu | Ser | Pro | Gln | Gly | Thr | Arg | Cys | Ile | Asp | Val | Asp | Glu | Cys | Arg |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Arg | Val | Pro | Pro | Pro | Cys | Ala | Pro | Gly | Arg | Cys | Glu | Asn | Ser | Pro |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Gly | Ser | Phe | Arg | Cys | Val | Cys | Gly | Pro | Gly | Phe | Arg | Ala | Gly | Pro |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Arg | Ala | Ala | Glu | Cys | Leu | Asp | Val | Asp | Glu | Cys | His | Arg | Val | Pro |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Pro | Pro | Cys | Asp | Leu | Gly | Arg | Cys | Glu | Asn | Thr | Pro | Gly | Ser | Phe |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Leu | Cys | Val | Cys | Pro | Ala | Gly | Tyr | Gln | Ala | Ala | Pro | His | Gly | Ala |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Ser | Cys | Gln | Asp | Val | Asp | Glu | Cys | Thr | Gln | Ser | Pro | Gly | Leu | Cys |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Gly | Arg | Gly | Ala | Cys | Lys | Asn | Leu | Pro | Gly | Ser | Phe | Arg | Cys | Val |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Cys | Pro | Ala | Gly | Phe | Arg | Gly | Ser | Ala | Cys | Glu | Glu | Asp | Val | Asp |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Glu | Cys | Ala | Gln | Glu | Pro | Pro | Pro | Cys | Gly | Pro | Gly | Arg | Cys | Asp |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Asn | Thr | Ala | Gly | Ser | Phe | His | Cys | Ala | Cys | Pro | Ala | Gly | Phe | Arg |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Ser | Arg | Gly | Pro | Gly | Ala | Pro | Cys | Gln | Asp | Val | Asp | Glu | Cys | Ala |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Arg | Ser | Pro | Pro | Pro | Cys | Thr | Tyr | Gly | Arg | Cys | Glu | Asn | Thr | Glu |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Gly | Ser | Phe | Gln | Cys | Val | Cys | Pro | Met | Gly | Phe | Gln | Pro | Asn | Thr |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Ala | Gly | Ser | Glu | Cys | Glu | Asp | Val | Asp | Glu | Cys | Glu | Asn | His | Leu |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Ala | Cys | Pro | Gly | Gln | Glu | Cys | Val | Asn | Ser | Pro | Gly | Ser | Phe | Gln |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Cys | Arg | Thr | Cys | Pro | Ser | Gly | His | His | Leu | His | Arg | Gly | Arg | Cys |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Thr | Asp | Val | Asp | Glu | Cys | Ser | Ser | Gly | Ala | Pro | Pro | Cys | Gly | Pro |
| | | | | 800 | | | | | 805 | | | | | 810 |
| His | Gly | His | Cys | Thr | Asn | Thr | Glu | Gly | Ser | Phe | Arg | Cys | Ser | Cys |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Ala | Pro | Gly | Tyr | Arg | Ala | Pro | Ser | Gly | Arg | Pro | Gly | Pro | Cys | Ala |
| | | | | 830 | | | | | 835 | | | | | 840 |
| Asp | Val | Asn | Glu | Cys | Leu | Glu | Gly | Asp | Phe | Cys | Phe | Pro | His | Gly |
| | | | | 845 | | | | | 850 | | | | | 855 |
| Glu | Cys | Leu | Asn | Thr | Asp | Gly | Ser | Phe | Ala | Cys | Thr | Cys | Ala | Pro |
| | | | | 860 | | | | | 865 | | | | | 870 |
| Gly | Tyr | Arg | Pro | Gly | Pro | Arg | Gly | Ala | Ser | Cys | Leu | Asp | Val | Asp |
| | | | | 875 | | | | | 880 | | | | | 885 |
| Glu | Cys | Ser | Glu | Glu | Asp | Leu | Cys | Gln | Ser | Gly | Ile | Cys | Thr | Asn |
| | | | | 890 | | | | | 895 | | | | | 900 |
| Thr | Asp | Gly | Ser | Phe | Glu | Cys | Ile | Cys | Pro | Pro | Gly | His | Arg | Ala |
| | | | | 905 | | | | | 910 | | | | | 915 |
| Gly | Pro | Asp | Leu | Ala | Ser | Cys | Leu | Asp | Val | Asp | Glu | Cys | Arg | Glu |
| | | | | 920 | | | | | 925 | | | | | 930 |
| Arg | Gly | Pro | Ala | Leu | Cys | Gly | Ser | Gln | Arg | Cys | Glu | Asn | Ser | Pro |

| | | | | | |
|---------------------|-----------------|---------------------|------|--|------|
| | 935 | | 940 | | 945 |
| Gly Ser Tyr Arg Cys | Val Arg Asp Cys | Asp Pro Gly Tyr His | Ala | | |
| | 950 | | 955 | | 960 |
| Gly Pro Glu Gly Thr | Cys Asp Asp Val | Asp Glu Cys Gln Glu | Tyr | | |
| | 965 | | 970 | | 975 |
| Gly Pro Glu Ile Cys | Gly Ala Gln Arg | Cys Glu Asn Thr Pro | Gly | | |
| | 980 | | 985 | | 990 |
| Ser Tyr Arg Cys Thr | Pro Ala Cys Asp | Pro Gly Tyr Gln Pro | Thr | | |
| | 995 | | 1000 | | 1005 |
| Pro Gly Gly Gly Cys | Gln Asp Val Asp | Glu Cys Arg Asn Arg | Ser | | |
| | 1010 | | 1015 | | 1020 |
| Phe Cys Gly Ala His | Ala Val Cys Gln | Asn Leu Pro Gly Ser | Phe | | |
| | 1025 | | 1030 | | 1035 |
| Gln Cys Leu Cys Asp | Gln Gly Tyr Glu | Gly Ala Arg Asp Gly | Arg | | |
| | 1040 | | 1045 | | 1050 |
| His Cys Val Asp Val | Asn Glu Cys Glu | Thr Leu Gln Gly Val | Cys | | |
| | 1055 | | 1060 | | 1065 |
| Gly Ala Ala Leu Cys | Glu Asn Val Glu | Gly Ser Phe Leu Cys | Val | | |
| | 1070 | | 1075 | | 1080 |
| Cys Pro Asn Ser Pro | Glu Glu Phe Asp | Pro Met Thr Gly Arg | Cys | | |
| | 1085 | | 1090 | | 1095 |
| Val Pro Pro Arg Thr | Ser Ala Asp Val | Asp Glu Cys Gln Leu | Phe | | |
| | 1100 | | 1105 | | 1110 |
| Arg Asp Gln Val Cys | Lys Ser Gly Val | Cys Val Asn Thr Ala | Pro | | |
| | 1115 | | 1120 | | 1125 |
| Gly Tyr Ser Cys Tyr | Cys Ser Asn Gly | Tyr Tyr Tyr His Thr | Gln | | |
| | 1130 | | 1135 | | 1140 |
| Arg Leu Glu Cys Ile | Asp Asn Asp Glu | Cys Ala Asp Glu Glu | Pro | | |
| | 1145 | | 1150 | | 1155 |
| Ala Cys Glu Gly Gly | Arg Cys Val Asn | Thr Val Gly Ser Tyr | His | | |
| | 1160 | | 1165 | | 1170 |
| Cys Thr Cys Glu Pro | Pro Leu Val Leu | Asp Gly Ser Gln Arg | Arg | | |
| | 1175 | | 1180 | | 1185 |
| Cys Val Ser Asn Glu | Ser Gln Ser Leu | Asp Asp Asn Leu Gly | Val | | |
| | 1190 | | 1195 | | 1200 |
| Cys Trp Gln Glu Val | Gly Ala Asp Leu | Val Cys Ser His Pro | Arg | | |
| | 1205 | | 1210 | | 1215 |
| Leu Asp Arg Gln Ala | Thr Tyr Thr Glu | Cys Cys Cys Leu Tyr | Gly | | |
| | 1220 | | 1225 | | 1230 |
| Glu Ala Trp Gly Met | Asp Cys Ala Leu | Cys Pro Ala Gln Asp | Ser | | |
| | 1235 | | 1240 | | 1245 |
| Asp Asp Phe Glu Ala | Leu Cys Asn Val | Leu Arg Pro Pro Ala | Tyr | | |
| | 1250 | | 1255 | | 1260 |
| Ser Pro Pro Arg Pro | Gly Gly Phe Gly | Leu Pro Tyr Glu Tyr | Gly | | |
| | 1265 | | 1270 | | 1275 |
| Pro Asp Leu Gly Pro | Pro Tyr Gln Gly | Leu Pro Tyr Gly Pro | Glu | | |
| | 1280 | | 1285 | | 1290 |
| Leu Tyr Pro Pro Pro | Ala Leu Pro Tyr | Asp Pro Tyr Pro Pro | Pro | | |
| | 1295 | | 1300 | | 1305 |
| Pro Gly Pro Phe Ala | Arg Arg Glu Ala | Pro Tyr Gly Ala Pro | Arg | | |
| | 1310 | | 1315 | | 1320 |
| Phe Asp Met Pro Asp | Phe Glu Asp Asp | Gly Gly Pro Tyr Gly | Glu | | |
| | 1325 | | 1330 | | 1335 |
| Ser Glu Ala Pro Ala | Pro Pro Gly Pro | Gly Thr Arg Trp Pro | Tyr | | |
| | 1340 | | 1345 | | 1350 |
| Arg Ser Arg Asp Thr | Arg Arg Ser Phe | Pro Glu Pro Glu Glu | Pro | | |
| | 1355 | | 1360 | | 1365 |
| Pro Glu Gly Gly Ser | Tyr Ala Gly Ser | Leu Ala Glu Pro Tyr | Glu | | |
| | 1370 | | 1375 | | 1380 |
| Glu Leu Glu Ala Glu | Glu Cys Gly Ile | Leu Asp Gly Cys Thr | Asn | | |
| | 1385 | | 1390 | | 1395 |
| Gly Arg Cys Val Arg | Val Pro Glu Gly | Phe Thr Cys Arg Cys | Phe | | |
| | 1400 | | 1405 | | 1410 |


```

Asp Gly Tyr Arg Leu Asp Met Thr Arg Met Ala Cys Val Asp Ile
      1415                      1420                      1425
Asn Glu Cys Asp Glu Ala Glu Ala Ala Ser Pro Leu Cys Val Asn
      1430                      1435                      1440
Ala Arg Cys Leu Asn Thr Asp Gly Ser Phe Arg Cys Ile Cys Arg
      1445                      1450                      1455
Pro Gly Phe Ala Pro Thr His Gln Pro His His Cys Ala Pro Ala
      1460                      1465                      1470
Arg Pro Arg Ala

```

```

<210> 5
<211> 273
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7505057CD1

```

```

<400> 5
Met Arg Ile Ala Val Ile Cys Phe Cys Leu Leu Gly Ile Thr Cys
  1           5           10           15
Ala Ile Pro Val Lys Gln Ala Asp Ser Gly Ser Ser Glu Glu Lys
      20           25           30
Gln Thr Leu Pro Ser Lys Ser Asn Glu Ser His Asp His Met Asp
      35           40           45
Asp Met Asp Asp Glu Asp Asp Asp Asp His Val Asp Ser Gln Asp
      50           55           60
Ser Ile Asp Ser Asn Asp Ser Asp Asp Val Asp Asp Thr Asp Asp
      65           70           75
Ser His Gln Ser Asp Glu Ser His His Ser Asp Glu Ser Asp Glu
      80           85           90
Leu Val Thr Asp Phe Pro Thr Asp Leu Pro Ala Thr Glu Val Phe
      95          100          105
Thr Pro Val Val Pro Thr Val Asp Thr Tyr Asp Gly Arg Gly Asp
     110          115          120
Ser Val Val Tyr Gly Leu Arg Ser Lys Ser Lys Lys Phe Arg Arg
     125          130          135
Pro Asp Ile Gln Tyr Pro Asp Ala Thr Asp Glu Asp Ile Thr Ser
     140          145          150
His Met Glu Ser Glu Glu Leu Asn Gly Ala Tyr Lys Ala Ile Pro
     155          160          165
Val Ala Gln Asp Leu Asn Ala Pro Ser Asp Trp Asp Ser Arg Gly
     170          175          180
Lys Asp Ser Tyr Glu Thr Ser Gln Leu Asp Asp Gln Ser Ala Glu
     185          190          195
Thr His Ser His Lys Gln Ser Arg Leu Tyr Lys Arg Lys Ala Asn
     200          205          210
Asp Glu Ser Asn Glu His Ser Asp Val Ile Asp Ser Gln Glu Leu
     215          220          225
Ser Lys Val Ser Arg Glu Phe His Ser His Glu Phe His Ser His
     230          235          240
Glu Asp Met Leu Val Val Asp Pro Lys Ser Lys Glu Glu Asp Lys
     245          250          255
His Leu Lys Phe Arg Ile Ser His Glu Leu Asp Ser Ala Ser Ser
     260          265          270
Glu Val Asn

```

```

<210> 6
<211> 333
<212> PRT

```

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 90116002CD1

<400> 6

| | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| Met | Arg | Ile | Trp | Trp | Leu | Leu | Leu | Ala | Ile | Glu | Ile | Cys | Thr | Gly | 1 | 5 | 10 | 15 |
| Asn | Ile | Asn | Ser | Gln | Asp | Thr | Cys | Arg | Gln | Gly | His | Pro | Gly | Ile | 20 | 25 | 30 | |
| Pro | Gly | Asn | Pro | Gly | His | Asn | Gly | Leu | Pro | Gly | Arg | Asp | Gly | Arg | 35 | 40 | 45 | |
| Asp | Gly | Ala | Lys | Gly | Asp | Lys | Gly | Asp | Ala | Gly | Glu | Pro | Gly | Arg | 50 | 55 | 60 | |
| Pro | Gly | Ser | Pro | Gly | Lys | Asp | Gly | Thr | Ser | Gly | Glu | Lys | Gly | Glu | 65 | 70 | 75 | |
| Arg | Gly | Ala | Asp | Gly | Lys | Val | Glu | Ala | Lys | Gly | Ile | Lys | Gly | Asp | 80 | 85 | 90 | |
| Gln | Gly | Ser | Arg | Gly | Ser | Pro | Gly | Lys | His | Gly | Pro | Lys | Gly | Leu | 95 | 100 | 105 | |
| Ala | Gly | Pro | Met | Gly | Glu | Lys | Gly | Leu | Arg | Gly | Glu | Thr | Gly | Pro | 110 | 115 | 120 | |
| Gln | Gly | Gln | Lys | Gly | Asn | Lys | Gly | Asp | Val | Gly | Pro | Thr | Gly | Pro | 125 | 130 | 135 | |
| Glu | Gly | Pro | Arg | Gly | Asn | Ile | Gly | Pro | Leu | Gly | Pro | Thr | Gly | Leu | 140 | 145 | 150 | |
| Pro | Gly | Pro | Met | Gly | Pro | Ile | Gly | Lys | Pro | Gly | Pro | Lys | Gly | Glu | 155 | 160 | 165 | |
| Ala | Gly | Pro | Thr | Gly | Pro | Gln | Gly | Glu | Pro | Gly | Val | Arg | Gly | Ile | 170 | 175 | 180 | |
| Arg | Gly | Trp | Lys | Gly | Asp | Arg | Gly | Glu | Lys | Gly | Lys | Ile | Gly | Glu | 185 | 190 | 195 | |
| Thr | Leu | Val | Leu | Pro | Lys | Ser | Ala | Phe | Thr | Val | Gly | Leu | Thr | Val | 200 | 205 | 210 | |
| Leu | Ser | Lys | Phe | Pro | Ser | Ser | Asp | Val | Pro | Ile | Lys | Phe | Asp | Lys | 215 | 220 | 225 | |
| Ile | Leu | Tyr | Asn | Glu | Phe | Asn | His | Tyr | Asp | Thr | Ala | Ala | Gly | Lys | 230 | 235 | 240 | |
| Phe | Thr | Cys | His | Ile | Ala | Gly | Val | Tyr | Tyr | Phe | Thr | Tyr | His | Ile | 245 | 250 | 255 | |
| Thr | Val | Phe | Ser | Arg | Asn | Val | Gln | Val | Ser | Leu | Val | Lys | Asn | Gly | 260 | 265 | 270 | |
| Val | Lys | Ile | Leu | His | Thr | Lys | Asp | Ala | Tyr | Met | Ser | Ser | Glu | Asp | 275 | 280 | 285 | |
| Gln | Ala | Ser | Gly | Gly | Ile | Val | Leu | Gln | Leu | Lys | Leu | Gly | Asp | Glu | 290 | 295 | 300 | |
| Val | Trp | Leu | Gln | Val | Thr | Gly | Gly | Glu | Arg | Phe | Asn | Gly | Leu | Phe | 305 | 310 | 315 | |
| Ala | Asp | Glu | Asp | Asp | Asp | Thr | Thr | Phe | Thr | Gly | Phe | Leu | Leu | Phe | 320 | 325 | 330 | |
| Ser | Ser | Pro | | | | | | | | | | | | | | | | |

<210> 7

<211> 478

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 039283CD1

<400> 7

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| Met | Glu | Thr | Leu | Lys | Asp | Lys | Thr | Leu | Gln | Glu | Leu | Glu | Glu | Leu | |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Gln | Asn | Asp | Ser | Glu | Ala | Ile | Asp | Gln | Leu | Ala | Leu | Glu | Ser | Pro | |
| | | | | 20 | | | | | 25 | | | | | 30 | |
| Glu | Val | Gln | Asp | Leu | Gln | Leu | Glu | Arg | Glu | Met | Ala | Leu | Ala | Thr | |
| | | | | 35 | | | | | 40 | | | | | 45 | |
| Asn | Arg | Ser | Leu | Ala | Glu | Arg | Asn | Leu | Glu | Phe | Gln | Gly | Pro | Leu | |
| | | | | 50 | | | | | 55 | | | | | 60 | |
| Glu | Ile | Ser | Arg | Ser | Asn | Leu | Ser | Asp | Arg | Tyr | Gln | Glu | Leu | Arg | |
| | | | | 65 | | | | | 70 | | | | | 75 | |
| Lys | Leu | Val | Glu | Arg | Cys | Gln | Glu | Gln | Lys | Ala | Lys | Leu | Glu | Lys | |
| | | | | 80 | | | | | 85 | | | | | 90 | |
| Phe | Ser | Ser | Ala | Leu | Gln | Pro | Gly | Thr | Leu | Leu | Asp | Leu | Leu | Gln | |
| | | | | 95 | | | | | 100 | | | | | 105 | |
| Val | Glu | Gly | Met | Lys | Ile | Glu | Glu | Glu | Ser | Glu | Ala | Met | Ala | Glu | |
| | | | | 110 | | | | | 115 | | | | | 120 | |
| Lys | Phe | Leu | Glu | Gly | Glu | Val | Pro | Leu | Glu | Thr | Phe | Leu | Glu | Asn | |
| | | | | 125 | | | | | 130 | | | | | 135 | |
| Phe | Ser | Ser | Met | Arg | Met | Leu | Ser | His | Leu | Arg | Arg | Val | Arg | Val | |
| | | | | 140 | | | | | 145 | | | | | 150 | |
| Glu | Lys | Leu | Gln | Glu | Val | Val | Arg | Lys | Pro | Arg | Ala | Ser | Gln | Glu | |
| | | | | 155 | | | | | 160 | | | | | 165 | |
| Leu | Ala | Gly | Asp | Ala | Pro | Pro | Pro | Arg | Pro | Pro | Pro | Pro | Val | Arg | |
| | | | | 170 | | | | | 175 | | | | | 180 | |
| Pro | Val | Pro | Gln | Gly | Thr | Pro | Pro | Val | Val | Glu | Glu | Gln | Pro | Gln | |
| | | | | 185 | | | | | 190 | | | | | 195 | |
| Pro | Pro | Ser | Ala | Met | Pro | Pro | Tyr | Pro | Leu | Pro | Tyr | Ser | Pro | Ser | |
| | | | | 200 | | | | | 205 | | | | | 210 | |
| Pro | Ser | Leu | Pro | Val | Gly | Pro | Thr | Ala | His | Gly | Ala | Leu | Pro | Pro | |
| | | | | 215 | | | | | 220 | | | | | 225 | |
| Ala | Pro | Phe | Pro | Val | Val | Ser | Gln | Pro | Ser | Phe | Tyr | Ser | Gly | Pro | |
| | | | | 230 | | | | | 235 | | | | | 240 | |
| Leu | Gly | Pro | Thr | Tyr | Pro | Ala | Ala | Gln | Leu | Gly | Pro | Arg | Gly | Ala | |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Ala | Gly | Tyr | Ser | Trp | Ser | Pro | Gln | Arg | Ser | Met | Pro | Pro | Arg | Pro | |
| | | | | 260 | | | | | 265 | | | | | 270 | |
| Gly | Tyr | Pro | Gly | Thr | Pro | Met | Gly | Ala | Ser | Gly | Pro | Gly | Tyr | Pro | |
| | | | | 275 | | | | | 280 | | | | | 285 | |
| Leu | Arg | Gly | Gly | Arg | Ala | Pro | Ser | Pro | Gly | Tyr | Pro | Gln | Gln | Ser | |
| | | | | 290 | | | | | 295 | | | | | 300 | |
| Pro | Tyr | Pro | Ala | Thr | Gly | Gly | Lys | Pro | Pro | Tyr | Pro | Ile | Gln | Pro | |
| | | | | 305 | | | | | 310 | | | | | 315 | |
| Gln | Leu | Pro | Ser | Phe | Pro | Gly | Gln | Pro | Gln | Pro | Ser | Val | Pro | Tyr | |
| | | | | 320 | | | | | 325 | | | | | 330 | |
| Ser | Pro | Ser | Pro | Ser | Leu | Ala | Val | Gly | Pro | Thr | Ala | His | Gly | Ala | |
| | | | | 335 | | | | | 340 | | | | | 345 | |
| Leu | Pro | Pro | Ala | Pro | Phe | Pro | Val | Val | Ser | Gln | Pro | Ser | Phe | Tyr | |
| | | | | 350 | | | | | 355 | | | | | 360 | |
| Ser | Gly | Pro | Leu | Gly | Pro | Thr | Tyr | Pro | Ala | Ala | Gln | Leu | Gly | Pro | |
| | | | | 365 | | | | | 370 | | | | | 375 | |
| Arg | Gly | Ala | Ala | Gly | Tyr | Ser | Trp | Ser | Pro | Gln | Arg | Ser | Met | Pro | |
| | | | | 380 | | | | | 385 | | | | | 390 | |
| Pro | Arg | Pro | Gly | Tyr | Pro | Gly | Thr | Pro | Met | Gly | Ala | Ser | Gly | Pro | |
| | | | | 395 | | | | | 400 | | | | | 405 | |
| Gly | Tyr | Pro | Leu | Arg | Gly | Gly | Arg | Ala | Pro | Ser | Pro | Gly | Tyr | Pro | |
| | | | | 410 | | | | | 415 | | | | | 420 | |
| Gln | Gln | Ser | Pro | Tyr | Pro | Ala | Thr | Gly | Gly | Lys | Pro | Pro | Tyr | Pro | |
| | | | | 425 | | | | | 430 | | | | | 435 | |
| Ile | Gln | Pro | Gln | Leu | Pro | Ser | Phe | Pro | Gly | Gln | Pro | Gln | Pro | Ser | |
| | | | | 440 | | | | | 445 | | | | | 450 | |
| Val | Pro | Leu | Gln | Pro | Pro | Tyr | Pro | Pro | Gly | Pro | Ala | Pro | Pro | Tyr | |
| | | | | 455 | | | | | 460 | | | | | 465 | |

Gly Phe Pro Pro Pro Pro Gly Pro Ala Trp Pro Gly Tyr
 470 475

<210> 8
 <211> 185
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7505082CD1

<400> 8
 Met Val Pro Glu Ala Trp Arg Ser Gly Leu Val Tyr Leu Thr Thr
 1 5 10 15
 Asp Ser Arg Arg Ser Asp Pro Leu Leu Lys Lys Pro Gly Ala Ala
 20 25 30
 Ser Pro Leu Ala Ser Arg Gln Asn Thr Leu Arg Ser Cys Asp Pro
 35 40 45
 Val Phe Tyr Arg Gln Val Leu Gly Ala Glu Ser Ala Pro Pro Gly
 50 55 60
 Gln Gln Ala Pro Pro Asn Thr Asp Trp Arg Phe Ser Gln Ala Gln
 65 70 75
 Arg Pro Gly Thr Ser Gly Ser Gln Asn Gly Asp Asp Thr Gly Thr
 80 85 90
 Trp Pro Asn Asn Gln Phe Asp Thr Glu Met Leu Gln Ala Met Ile
 95 100 105
 Leu Ala Ser Ala Ser Glu Ala Ala Asp Gly Ser Ser Thr Leu Gly
 110 115 120
 Gly Gly Ala Gly Thr Met Gly Leu Ser Ala Arg Tyr Gly Pro Gln
 125 130 135
 Phe Thr Leu Gln His Val Pro Asp Tyr Arg Gln Asn Val Tyr Ile
 140 145 150
 Pro Gly Ser Asn Ala Thr Leu Thr Asn Ala Ala Gly Lys Arg Asp
 155 160 165
 Gly Lys Ala Pro Ala Gly Gly Asn Gly Asn Lys Lys Lys Ser Gly
 170 175 180
 Lys Lys Glu Lys Lys
 185

<210> 9
 <211> 170
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7505139CD1

<400> 9
 Met Ser Arg Leu Ser Arg Ser Leu Leu Trp Ala Ala Thr Cys Leu
 1 5 10 15
 Gly Val Leu Cys Val Leu Ser Ala Asp Lys Asn Thr Thr Gln His
 20 25 30
 Pro Asn Val Thr Thr Leu Ala Pro Ile Ser Asn Val Thr Ser Ala
 35 40 45
 Pro Val Thr Ser Leu Pro Leu Val Thr Thr Pro Ala Pro Glu Thr
 50 55 60
 Cys Glu Gly Arg Asn Ser Cys Val Ser Cys Phe Asn Val Ser Val
 65 70 75
 Val Asn Thr Thr Cys Phe Trp Ile Glu Cys Lys Asp Glu Ser Tyr
 80 85 90
 Cys Ser His Asn Ser Thr Val Ser Asp Cys Gln Val Gly Asn Thr

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | 95 | | | | | 100 | | | | | 105 |
| Thr | Asp | Phe | Cys | Ser | Val | Ser | Thr | Ala | Thr | Pro | Val | Pro | Thr | Ala |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Asn | Ser | Thr | Gly | Thr | Thr | Asn | Asn | Thr | Val | Thr | Pro | Thr | Ser | Gln |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Pro | Val | Arg | Lys | Ser | Thr | Phe | Asp | Ala | Ala | Ser | Phe | Ile | Gly | Gly |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Ile | Val | Leu | Val | Leu | Glu | Ile | Arg | Cys | His | Thr | Arg | Asn | Tyr | Ile |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Pro | Asp | Leu | Lys | Lys | | | | | | | | | | |
| | | | | 170 | | | | | | | | | | |

<210> 10
 <211> 482
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7505234CD1

<400> 10

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Ala | Gly | Glu | Gly | Lys | Glu | Arg | Val | Pro | Lys | Gln | Arg | Gln |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Val | Leu | Ile | Phe | Phe | Val | Leu | Leu | Gly | Ile | Ala | Gln | Ala | Ser | Cys |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Gln | Pro | Arg | His | Tyr | Ser | Val | Ala | Glu | Glu | Thr | Glu | Ser | Gly | Ser |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Phe | Val | Ala | Asn | Leu | Leu | Lys | Asp | Leu | Gly | Leu | Glu | Ile | Gly | Glu |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Leu | Ala | Val | Arg | Gly | Ala | Arg | Val | Val | Ser | Lys | Gly | Lys | Lys | Met |
| | | | | 65 | | | | | 70 | | | | | 75 |
| His | Leu | Gln | Phe | Asp | Arg | Gln | Thr | Gly | Asp | Leu | Leu | Leu | Asn | Glu |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Lys | Leu | Asp | Arg | Glu | Glu | Leu | Cys | Gly | Pro | Thr | Glu | Pro | Cys | Val |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Leu | Pro | Phe | Gln | Val | Leu | Leu | Glu | Asn | Pro | Leu | Gln | Phe | Phe | Gln |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Ala | Glu | Leu | Arg | Ile | Arg | Asp | Val | Asn | Asp | His | Ser | Pro | Val | Phe |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Leu | Asp | Lys | Glu | Ile | Leu | Leu | Lys | Ile | Pro | Glu | Ser | Ile | Thr | Pro |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Gly | Thr | Thr | Phe | Leu | Ile | Glu | Arg | Ala | Gln | Asp | Leu | Asp | Val | Gly |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Thr | Asn | Ser | Leu | Gln | Asn | Tyr | Thr | Ile | Ser | Pro | Asn | Phe | His | Phe |
| | | | | 170 | | | | | 175 | | | | | 180 |
| His | Leu | Asn | Leu | Gln | Asp | Ser | Leu | Asp | Gly | Ile | Ile | Leu | Pro | Gln |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Leu | Val | Leu | Asn | Arg | Ala | Leu | Asp | Arg | Glu | Glu | Gln | Pro | Glu | Ile |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Arg | Leu | Thr | Leu | Thr | Ala | Leu | Asp | Gly | Gly | Ser | Pro | Pro | Arg | Ser |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Gly | Thr | Ala | Leu | Val | Arg | Ile | Glu | Val | Val | Asp | Ile | Asn | Asp | Asn |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Val | Pro | Glu | Phe | Ala | Lys | Leu | Leu | Tyr | Glu | Val | Gln | Ile | Pro | Glu |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Asp | Ser | Pro | Val | Gly | Ser | Gln | Val | Ala | Ile | Val | Ser | Ala | Arg | Asp |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Leu | Asp | Ile | Gly | Thr | Asn | Gly | Glu | Ile | Ser | Tyr | Ala | Phe | Ser | Gln |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Ala | Ser | Glu | Asp | Ile | Arg | Lys | Thr | Phe | Arg | Leu | Ser | Ala | Lys | Ser |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Gly | Glu | Leu | Leu | Leu | Arg | Gln | Lys | Leu | Asp | Phe | Glu | Ser | Ile | Gln |

| | | | | | |
|-----------------|-----|---------------------|-----|-------------------------|-----|
| Thr Tyr Thr Val | 305 | Asn Ile Gln Ala Thr | 310 | Asp Gly Gly Gly Leu Ser | 315 |
| | 320 | Val Phe Val Gln Val | 325 | Met Asp Leu Asn Asp Asn | 330 |
| Gly Thr Cys Val | 335 | | 340 | | 345 |
| Pro Pro Glu Leu | 350 | Thr Met Ser Thr Leu | 355 | Ile Asn Gln Ile Pro Glu | 360 |
| Asn Leu Gln Asp | 365 | Thr Leu Ile Ala Val | 370 | Phe Ser Val Ser Asp Pro | 375 |
| Asp Ser Gly Asp | 380 | Asn Gly Arg Met Val | 385 | Cys Ser Ile Gln Asp Asp | 390 |
| Leu Pro Phe Phe | 395 | Leu Lys Pro Ser Val | 400 | Glu Asn Phe Tyr Thr Leu | 405 |
| Val Ile Ser Thr | 410 | Ala Leu Asp Arg Glu | 415 | Thr Arg Ser Glu Tyr Asn | 420 |
| Ile Thr Ile Thr | 425 | Val Thr Asp Phe Gly | 430 | Thr Leu Ser Gln Ser Tyr | 435 |
| Gln Tyr Glu Val | 440 | Cys Leu Thr Gly Gly | 445 | Ser Gly Thr Asn Glu Phe | 450 |
| Lys Phe Leu Lys | 455 | Pro Ile Ile Pro Asn | 460 | Phe Val Ala Gln Gly Ala | 465 |
| Glu Arg Val Ser | 470 | Glu Ala Asn Pro Ser | 475 | Phe Arg Lys Ser Phe Glu | 480 |
| Phe Thr | | | | | |

<210> 11
 <211> 117
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7500227CD1

| | | | | | |
|---------------------|-----|---------------------|-----|---------------------|-----|
| Met Cys Ser Arg Val | 1 | Pro Leu Leu Leu Pro | 10 | Leu Leu Leu Leu Leu | 15 |
| Ala Leu Gly Pro Gly | 20 | Val Gln Gly Cys Pro | 25 | Ser Gly Cys Gln Cys | 30 |
| Ser Gln Pro Gln Thr | 35 | Val Phe Cys Thr Ala | 40 | Arg Gln Gly Thr Thr | 45 |
| Val Pro Arg Asp Val | 50 | Pro Pro Asp Thr Val | 55 | Gly Leu Tyr Val Phe | 60 |
| Glu Asn Gly Ile Thr | 65 | Met Leu Asp Ala Gly | 70 | Ser Phe Ala Gly Leu | 75 |
| Pro Gly Leu Gln Leu | 80 | Leu Asp Leu Ser Gln | 85 | Asn Gln Ile Ala Ser | 90 |
| Leu Pro Ser Gly Val | 95 | Phe Gln Pro Leu Met | 100 | Gly Phe Pro Gly Pro | 105 |
| Gly Leu Gln Ser Pro | 110 | Leu His Ala Lys Pro | 115 | Tyr Ile | |

<210> 12
 <211> 485
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503676CD1

<400> 12

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Ser | Arg | Ala | Val | Val | Arg | Ala | Arg | Arg | Cys | Pro | Gln | Cys |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Pro | Gln | Val | Arg | Ala | Ala | Ala | Ala | Ala | Pro | Ala | Trp | Ala | Ala | Leu |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Pro | Leu | Ser | Arg | Ser | Leu | Pro | Pro | Cys | Ser | Asn | Ser | Ser | Ser | Phe |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Ser | Met | Pro | Leu | Phe | Leu | Leu | Leu | Leu | Leu | Val | Leu | Leu | Leu | Leu |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Leu | Glu | Asp | Ala | Gly | Ala | Gln | Gln | Gly | Lys | Tyr | Cys | Gly | Leu | Gly |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Leu | Gln | Met | Asn | His | Ser | Ile | Glu | Ser | Lys | Gly | Asn | Glu | Ile | Thr |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Leu | Leu | Phe | Met | Ser | Gly | Ile | His | Val | Ser | Gly | Arg | Gly | Phe | Leu |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Ala | Ser | Tyr | Ser | Val | Ile | Asp | Lys | Gln | Asp | Leu | Ile | Thr | Cys | Leu |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Asp | Thr | Ala | Ser | Asn | Phe | Leu | Glu | Pro | Glu | Phe | Ser | Lys | Tyr | Cys |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Pro | Ala | Gly | Cys | Leu | Leu | Pro | Phe | Ala | Glu | Ile | Ser | Gly | Thr | Ile |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Pro | His | Gly | Tyr | Arg | Asp | Ser | Ser | Pro | Leu | Cys | Met | Ala | Gly | Val |
| | | | | 155 | | | | | 160 | | | | | 165 |
| His | Ala | Gly | Val | Val | Ser | Asn | Thr | Leu | Gly | Gly | Gln | Ile | Ser | Val |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Val | Ile | Ser | Lys | Gly | Ile | Pro | Tyr | Tyr | Glu | Ser | Ser | Leu | Ala | Asn |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Asn | Val | Thr | Ser | Val | Val | Gly | His | Leu | Ser | Thr | Ser | Leu | Phe | Thr |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Phe | Lys | Thr | Ser | Gly | Cys | Tyr | Gly | Thr | Leu | Gly | Met | Glu | Ser | Gly |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Val | Ile | Ala | Asp | Pro | Gln | Ile | Thr | Ala | Ser | Ser | Val | Leu | Glu | Trp |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Thr | Asp | His | Thr | Gly | Gln | Glu | Asn | Ser | Trp | Lys | Pro | Lys | Lys | Ala |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Arg | Leu | Lys | Lys | Pro | Gly | Pro | Pro | Trp | Ala | Ala | Phe | Ala | Thr | Asp |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Glu | Tyr | Gln | Trp | Leu | Gln | Ile | Asp | Leu | Asn | Lys | Glu | Lys | Lys | Ile |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Thr | Gly | Ile | Ile | Thr | Thr | Gly | Ser | Thr | Met | Val | Glu | His | Asn | Tyr |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Tyr | Val | Ser | Ala | Tyr | Arg | Ile | Leu | Tyr | Ser | Asp | Asp | Gly | Gln | Lys |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Trp | Thr | Val | Tyr | Arg | Glu | Pro | Gly | Val | Glu | Gln | Asp | Lys | Ile | Phe |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Gln | Gly | Asn | Lys | Asp | Tyr | His | Gln | Asp | Val | Arg | Asn | Asn | Phe | Leu |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Pro | Pro | Ile | Ile | Ala | Arg | Phe | Ile | Arg | Val | Asn | Pro | Thr | Gln | Trp |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Gln | Gln | Lys | Ile | Ala | Met | Lys | Met | Glu | Leu | Leu | Gly | Cys | Gln | Phe |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Ile | Pro | Lys | Gly | Arg | Pro | Pro | Lys | Leu | Thr | Gln | Pro | Pro | Pro | Pro |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Arg | Asn | Ser | Asn | Asp | Leu | Lys | Asn | Thr | Thr | Ala | Pro | Pro | Lys | Ile |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Ala | Lys | Gly | Arg | Ala | Pro | Lys | Phe | Thr | Gln | Pro | Leu | Gln | Pro | Arg |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Ser | Ser | Asn | Glu | Phe | Pro | Ala | Gln | Thr | Glu | Gln | Thr | Thr | Ala | Ser |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Pro | Asp | Ile | Arg | Asn | Thr | Thr | Val | Thr | Pro | Asn | Val | Thr | Lys | Asp |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Val | Ala | Leu | Ala | Ala | Val | Leu | Val | Pro | Val | Leu | Val | Met | Val | Leu |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Thr | Thr | Leu | Ile | Leu | Ile | Leu | Val | Cys | Ala | Trp | His | Trp | Arg | Asn |

470
 Arg Leu Val His Asn
 485

<210> 13
 <211> 1489
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503606CD1

<400> 13
 Met Gly Lys Glu Gln Glu Leu Val Gln Ala Val Lys Ala Glu Asp
 1 5 10 15
 Val Gly Thr Ala Gln Arg Leu Leu Gln Arg Pro Arg Pro Gly Lys
 20 25 30
 Ala Thr Arg Ser Leu Pro Gly Gly Arg Arg Trp Met Asp Gly
 35 40 45
 Arg Val Asp Gln Pro Arg Val Arg Leu Arg Thr Tyr Ser Arg Val
 50 55 60
 Ser Val Ser Gly His Leu Cys Gly His Gly Gln Gly Ser Ala Glu
 65 70 75
 Leu Leu Gly Ser Thr Lys Lys Ile Asn Val Asn Phe Gln Asp Pro
 80 85 90
 Asp Gly Val Gly Phe Gly Val Lys Gly Gln Leu Pro Ala Ser Pro
 95 100 105
 Arg Pro Pro Gly Met Arg Pro Leu His Tyr Ala Ala Trp Gln Gly
 110 115 120
 Arg Lys Glu Pro Met Lys Leu Val Leu Lys Ala Gly Ser Ala Val
 125 130 135
 Asn Ile Pro Ser Asp Glu Gly His Ile Pro Leu His Leu Ala Ala
 140 145 150
 Gln His Gly His Tyr Asp Val Ser Glu Met Leu Leu Gln His Gln
 155 160 165
 Ser Asn Pro Cys Met Val Asp Asn Ser Gly Lys Thr Pro Leu Asp
 170 175 180
 Leu Ala Cys Glu Phe Gly Arg Val Gly Val Val Gln Leu Leu Leu
 185 190 195
 Ser Ser Asn Met Cys Ala Ala Leu Leu Glu Pro Arg Pro Gly Asp
 200 205 210
 Ala Thr Asp Pro Asn Gly Thr Ser Pro Leu His Leu Ala Ala Lys
 215 220 225
 Asn Gly His Ile Asp Ile Ile Arg Leu Leu Leu Gln Ala Gly Ile
 230 235 240
 Asp Ile Asn Arg Gln Thr Lys Ser Gly Thr Ala Leu His Glu Ala
 245 250 255
 Ala Leu Cys Gly Lys Thr Glu Val Val Arg Leu Leu Leu Asp Ser
 260 265 270
 Gly Ile Asn Ala His Val Arg Asn Thr Tyr Ser Gln Thr Ala Leu
 275 280 285
 Asp Ile Val His Gln Phe Thr Thr Ser Gln Ala Ser Arg Glu Ile
 290 295 300
 Lys Gln Leu Leu Arg Gly Gly Pro Arg Gln Gly Arg Gly Ala Val
 305 310 315
 Ala Gly Ala Gly Ala Gly Pro Asp Gln Pro Ala Gly Leu Arg Gly
 320 325 330
 Pro Pro Pro Val Ser Pro Glu Ala Ser Ala Ala Leu Gln Val Arg
 335 340 345
 Ala Thr Lys Asp Tyr Cys Asn Asn Tyr Asp Leu Thr Ser Leu Asn
 350 355 360
 Val Lys Ala Gly Asp Ile Ile Thr Val Leu Glu Gln His Pro Asp

| | | | | | |
|---------------------|---------------------|---------------------|-----|--|-----|
| | 365 | | 370 | | 375 |
| Gly Arg Trp Lys | Gly Cys Ile His Asp | Arg Thr Gly Asn Asp | | | |
| | 380 | | 385 | | 390 |
| Arg Val Gly Tyr Phe | Pro Ser Ser Leu Gly | Glu Ala Ile Val Lys | | | |
| | 395 | | 400 | | 405 |
| Arg Ala Gly Ser Arg | Ala Gly Thr Glu Pro | Ser Leu Pro Gln Gly | | | |
| | 410 | | 415 | | 420 |
| Ser Ser Ser Ser Gly | Pro Ser Ala Pro Pro | Glu Glu Ile Trp Val | | | |
| | 425 | | 430 | | 435 |
| Leu Arg Lys Pro Phe | Ala Gly Gly Asp Arg | Ser Gly Ser Ile Ser | | | |
| | 440 | | 445 | | 450 |
| Gly Met Ala Gly Gly | Arg Gly Ser Gly Gly | His Ala Leu His Ala | | | |
| | 455 | | 460 | | 465 |
| Gly Ser Glu Gly Val | Lys Leu Leu Ala Thr | Val Leu Ser Gln Lys | | | |
| | 470 | | 475 | | 480 |
| Ser Val Ser Glu Ser | Gly Pro Gly Asp Ser | Pro Ala Lys Pro Pro | | | |
| | 485 | | 490 | | 495 |
| Glu Gly Ser Ala Gly | Val Ala Arg Ser Gln | Pro Pro Val Ala His | | | |
| | 500 | | 505 | | 510 |
| Ala Gly Gln Val Tyr | Gly Glu Gln Pro Pro | Lys Lys Leu Glu Pro | | | |
| | 515 | | 520 | | 525 |
| Ala Ser Glu Gly Lys | Ser Ser Glu Ala Val | Ser Gln Trp Leu Thr | | | |
| | 530 | | 535 | | 540 |
| Ala Phe Gln Leu Gln | Leu Tyr Ala Pro Asn | Phe Ile Ser Ala Gly | | | |
| | 545 | | 550 | | 555 |
| Tyr Asp Leu Pro Thr | Ile Ser Arg Met Thr | Pro Glu Asp Leu Thr | | | |
| | 560 | | 565 | | 570 |
| Ala Ile Gly Val Thr | Lys Pro Gly His Arg | Lys Lys Ile Ala Ala | | | |
| | 575 | | 580 | | 585 |
| Glu Ile Ser Gly Leu | Ser Ile Pro Asp Trp | Leu Pro Glu His Lys | | | |
| | 590 | | 595 | | 600 |
| Pro Ala Asn Leu Ala | Val Trp Leu Ser Met | Ile Gly Leu Ala Gln | | | |
| | 605 | | 610 | | 615 |
| Tyr Tyr Lys Val Leu | Val Asp Asn Gly Tyr | Glu Asn Ile Asp Phe | | | |
| | 620 | | 625 | | 630 |
| Ile Thr Asp Ile Thr | Trp Glu Asp Leu Gln | Glu Ile Gly Ile Thr | | | |
| | 635 | | 640 | | 645 |
| Lys Leu Gly His Gln | Lys Lys Leu Met Leu | Ala Val Arg Lys Leu | | | |
| | 650 | | 655 | | 660 |
| Ala Glu Leu Gln Lys | Ala Glu Tyr Ala Lys | Tyr Glu Gly Gly Pro | | | |
| | 665 | | 670 | | 675 |
| Leu Arg Arg Lys Ala | Pro Gln Ser Leu Glu | Val Met Ala Ile Glu | | | |
| | 680 | | 685 | | 690 |
| Ser Pro Pro Pro Pro | Glu Pro Thr Pro Ala | Asp Cys Gln Ser Pro | | | |
| | 695 | | 700 | | 705 |
| Lys Met Thr Thr Phe | Gln Asp Ser Glu Leu | Ser Asp Glu Leu Gln | | | |
| | 710 | | 715 | | 720 |
| Ala Ala Met Thr Gly | Pro Ala Glu Val Gly | Pro Thr Thr Glu Lys | | | |
| | 725 | | 730 | | 735 |
| Pro Ser Ser His Leu | Pro Pro Thr Pro Arg | Ala Thr Thr Arg Gln | | | |
| | 740 | | 745 | | 750 |
| Asp Ser Ser Leu Gly | Gly Arg Ala Arg His | Met Ser Ser Ser Gln | | | |
| | 755 | | 760 | | 765 |
| Glu Leu Leu Gly Asp | Gly Pro Pro Gly Pro | Ser Ser Pro Met Ser | | | |
| | 770 | | 775 | | 780 |
| Arg Ser Gln Glu Tyr | Leu Leu Asp Glu Gly | Pro Ala Pro Gly Thr | | | |
| | 785 | | 790 | | 795 |
| Pro Pro Arg Glu Ala | Arg Pro Gly Arg His | Gly His Ser Ile Lys | | | |
| | 800 | | 805 | | 810 |
| Arg Ala Ser Val Pro | Pro Val Pro Gly Lys | Pro Arg Gln Val Leu | | | |
| | 815 | | 820 | | 825 |
| Pro Pro Gly Thr Ser | His Phe Thr Pro Pro | Gln Thr Pro Thr Lys | | | |
| | 830 | | 835 | | 840 |

| | | | | | | | | | | | | |
|-----------------|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|--|
| Thr Arg Pro Gly | Ser | Pro | Gln | Ala | Leu | Gly | Gly | Pro | His | Gly | Pro | |
| | 845 | | | | | 850 | | | | | 855 | |
| Ala Pro Ala Thr | Ala | Lys | Val | Lys | Pro | Thr | Pro | Gln | Leu | Leu | Pro | |
| | 860 | | | | | 865 | | | | | 870 | |
| Pro Thr Glu Arg | Pro | Met | Ser | Pro | Arg | Ser | Leu | Pro | Gln | Ser | Pro | |
| | 875 | | | | | 880 | | | | | 885 | |
| Thr His Arg Gly | Phe | Ala | Tyr | Val | Leu | Pro | Gln | Pro | Val | Glu | Gly | |
| | 890 | | | | | 895 | | | | | 900 | |
| Glu Val Gly Pro | Ala | Ala | Pro | Gly | Pro | Ala | Pro | Pro | Pro | Val | Pro | |
| | 905 | | | | | 910 | | | | | 915 | |
| Thr Ala Val Pro | Thr | Leu | Cys | Leu | Pro | Pro | Glu | Ala | Asp | Ala | Glu | |
| | 920 | | | | | 925 | | | | | 930 | |
| Pro Gly Arg Pro | Lys | Lys | Arg | Ala | His | Ser | Leu | Asn | Arg | Tyr | Ala | |
| | 935 | | | | | 940 | | | | | 945 | |
| Ala Ser Asp Ser | Glu | Pro | Glu | Arg | Asp | Glu | Leu | Leu | Val | Pro | Ala | |
| | 950 | | | | | 955 | | | | | 960 | |
| Ala Ala Gly Pro | Tyr | Ala | Thr | Val | Gln | Arg | Arg | Val | Gly | Arg | Ser | |
| | 965 | | | | | 970 | | | | | 975 | |
| His Ser Val Arg | Ala | Pro | Ala | Gly | Ala | Asp | Lys | Asn | Val | Asn | Arg | |
| | 980 | | | | | 985 | | | | | 990 | |
| Ser Gln Ser Phe | Ala | Val | Arg | Pro | Arg | Lys | Lys | Gly | Pro | Pro | Pro | |
| | 995 | | | | | 1000 | | | | | 1005 | |
| Pro Pro Pro Lys | Arg | Ser | Ser | Ser | Ala | Leu | Ala | Ser | Ala | Asn | Leu | |
| | 1010 | | | | | 1015 | | | | | 1020 | |
| Ala Asp Glu Pro | Val | Pro | Asp | Ala | Glu | Pro | Glu | Asp | Gly | Leu | Leu | |
| | 1025 | | | | | 1030 | | | | | 1035 | |
| Gly Val Arg Ala | Gln | Cys | Arg | Arg | Ala | Ser | Asp | Leu | Ala | Gly | Ser | |
| | 1040 | | | | | 1045 | | | | | 1050 | |
| Val Asp Thr Gly | Ser | Ala | Gly | Ser | Val | Lys | Ser | Ile | Ala | Ala | Met | |
| | 1055 | | | | | 1060 | | | | | 1065 | |
| Leu Glu Leu Ser | Ser | Ile | Gly | Gly | Gly | Gly | Arg | Ala | Ala | Arg | Arg | |
| | 1070 | | | | | 1075 | | | | | 1080 | |
| Pro Pro Glu Gly | His | Pro | Thr | Pro | Arg | Pro | Ala | Ser | Pro | Glu | Pro | |
| | 1085 | | | | | 1090 | | | | | 1095 | |
| Gly Arg Val Ala | Thr | Val | Leu | Ala | Ser | Val | Lys | His | Lys | Glu | Ala | |
| | 1100 | | | | | 1105 | | | | | 1110 | |
| Ile Gly Pro Gly | Gly | Glu | Val | Val | Asn | Arg | Arg | Arg | Thr | Leu | Ser | |
| | 1115 | | | | | 1120 | | | | | 1125 | |
| Gly Pro Val Thr | Gly | Leu | Leu | Ala | Thr | Ala | Arg | Arg | Gly | Pro | Gly | |
| | 1130 | | | | | 1135 | | | | | 1140 | |
| Glu Ser Ala Asp | Pro | Gly | Pro | Phe | Val | Glu | Asp | Gly | Thr | Gly | Arg | |
| | 1145 | | | | | 1150 | | | | | 1155 | |
| Gln Arg Pro Arg | Gly | Pro | Ser | Lys | Gly | Glu | Ala | Gly | Val | Glu | Gly | |
| | 1160 | | | | | 1165 | | | | | 1170 | |
| Pro Pro Leu Ala | Lys | Val | Glu | Ala | Ser | Ala | Thr | Leu | Lys | Arg | Arg | |
| | 1175 | | | | | 1180 | | | | | 1185 | |
| Ile Arg Ala Lys | Gln | Asn | Gln | Gln | Glu | Asn | Val | Lys | Phe | Ile | Leu | |
| | 1190 | | | | | 1195 | | | | | 1200 | |
| Thr Glu Ser Asp | Thr | Val | Lys | Arg | Arg | Pro | Lys | Ala | Lys | Glu | Arg | |
| | 1205 | | | | | 1210 | | | | | 1215 | |
| Glu Ala Gly Pro | Glu | Pro | Pro | Pro | Pro | Leu | Ser | Val | Tyr | His | Asn | |
| | 1220 | | | | | 1225 | | | | | 1230 | |
| Gly Thr Gly Thr | Val | Arg | Arg | Arg | Pro | Ala | Ser | Glu | Gln | Ala | Gly | |
| | 1235 | | | | | 1240 | | | | | 1245 | |
| Pro Pro Glu Leu | Pro | Pro | Pro | Pro | Pro | Pro | Ala | Glu | Pro | Pro | Pro | |
| | 1250 | | | | | 1255 | | | | | 1260 | |
| Thr Asp Leu Ala | His | Leu | Pro | Pro | Leu | Pro | Pro | Pro | Glu | Gly | Glu | |
| | 1265 | | | | | 1270 | | | | | 1275 | |
| Ala Arg Lys Pro | Ala | Lys | Pro | Pro | Val | Ser | Pro | Lys | Pro | Val | Leu | |
| | 1280 | | | | | 1285 | | | | | 1290 | |
| Thr Gln Pro Val | Pro | Lys | Leu | Gln | Gly | Ser | Pro | Thr | Pro | Thr | Ser | |
| | 1295 | | | | | 1300 | | | | | 1305 | |
| Lys Lys Val Pro | Leu | Pro | Gly | Pro | Gly | Ser | Pro | Glu | Val | Lys | Arg | |

| | | |
|---|------|------|
| 1310 | 1315 | 1320 |
| Ala His Gly Thr Pro Pro Val Ser Pro Lys Pro Pro Pro Pro | | |
| 1325 | 1330 | 1335 |
| Pro Thr Ala Pro Lys Pro Val Lys Ala Val Ala Gly Leu Pro Ser | | |
| 1340 | 1345 | 1350 |
| Gly Ser Ala Gly Pro Ser Pro Ala Pro Ser Pro Ala Arg Gln Pro | | |
| 1355 | 1360 | 1365 |
| Pro Ala Ala Leu Ala Lys Pro Pro Gly Thr Pro Pro Ser Leu Gly | | |
| 1370 | 1375 | 1380 |
| Ala Ser Pro Ala Lys Pro Pro Ser Pro Gly Ala Pro Ala Leu His | | |
| 1385 | 1390 | 1395 |
| Val Pro Ala Lys Pro Pro Arg Ala Ala Ala Ala Ala Ala Ala | | |
| 1400 | 1405 | 1410 |
| Ala Ala Ala Pro Pro Ala Pro Pro Glu Gly Ala Ser Pro Gly Asp | | |
| 1415 | 1420 | 1425 |
| Ser Ala Arg Gln Lys Leu Glu Glu Thr Ser Ala Cys Leu Ala Ala | | |
| 1430 | 1435 | 1440 |
| Ala Leu Gln Ala Val Glu Glu Lys Ile Arg Gln Glu Asp Ala Gln | | |
| 1445 | 1450 | 1455 |
| Gly Pro Arg Asp Ser Ala Ala Glu Lys Ser Thr Gly Ser Ile Leu | | |
| 1460 | 1465 | 1470 |
| Asp Asp Ile Gly Ser Met Phe Asp Asp Leu Ala Asp Gln Leu Asp | | |
| 1475 | 1480 | 1485 |
| Ala Met Leu Glu | | |

<210> 14

<211> 703

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500216CD1

<400> 14

| | | |
|---|-----|-----|
| Met Gly Ser Trp Ala Leu Leu Trp Pro Pro Leu Leu Phe Thr Gly | | |
| 1 | 5 | 10 |
| 15 | | |
| Leu Leu Val Arg Pro Pro Gly Thr Met Ala Gln Ala Gln Tyr Cys | | |
| 20 | 25 | 30 |
| Ser Val Asn Lys Asp Ile Phe Glu Val Glu Glu Asn Thr Asn Val | | |
| 35 | 40 | 45 |
| Thr Glu Pro Leu Val Asp Ile His Val Pro Glu Gly Gln Glu Val | | |
| 50 | 55 | 60 |
| Thr Leu Gly Ala Leu Ser Thr Pro Phe Ala Phe Arg Ile Gln Gly | | |
| 65 | 70 | 75 |
| Asn Gln Leu Phe Leu Asn Val Thr Pro Asp Tyr Glu Glu Lys Ser | | |
| 80 | 85 | 90 |
| Leu Leu Glu Ala Gln Leu Leu Cys Gln Ser Gly Gly Thr Leu Val | | |
| 95 | 100 | 105 |
| Thr Gln Leu Arg Val Phe Val Ser Val Leu Asp Val Asn Asp Asn | | |
| 110 | 115 | 120 |
| Ala Pro Glu Phe Pro Phe Lys Thr Lys Glu Ile Leu Val Glu Glu | | |
| 125 | 130 | 135 |
| Asp Thr Lys Val Asn Ser Thr Val Ile Pro Glu Thr Gln Leu Gln | | |
| 140 | 145 | 150 |
| Ala Glu Asp Arg Asp Lys Asp Asp Ile Leu Phe Tyr Thr Leu Gln | | |
| 155 | 160 | 165 |
| Glu Met Thr Ala Gly Ala Ser Asp Tyr Phe Ser Leu Val Ser Val | | |
| 170 | 175 | 180 |
| Asn Arg Pro Ala Leu Arg Leu Asp Arg Pro Leu Asp Phe Tyr Glu | | |
| 185 | 190 | 195 |
| Arg Pro Asn Met Thr Phe Trp Leu Leu Val Arg Asp Thr Pro Gly | | |

| | | | | | |
|-----------------|---------------------|---------------------|-----|--|-----|
| | 200 | | 205 | | 210 |
| Glu Asn Val Glu | Pro Ser His Thr Ala | Thr Ala Thr Leu Val | Leu | | |
| | 215 | | 220 | | 225 |
| Asn Val Val Pro | Ala Asp Leu Arg Pro | Pro Trp Phe Leu Pro | Cys | | |
| | 230 | | 235 | | 240 |
| Thr Phe Ser Asp | Gly Tyr Val Cys Ile | Gln Ala Gln Tyr His | Gly | | |
| | 245 | | 250 | | 255 |
| Ala Val Pro Thr | Gly His Ile Leu Pro | Ser Pro Leu Val Leu | Arg | | |
| | 260 | | 265 | | 270 |
| Pro Gly Pro Ile | Tyr Ala Glu Asp Gly | Asp Arg Gly Ile Asn | Gln | | |
| | 275 | | 280 | | 285 |
| Pro Ile Ile Tyr | Ser Ile Phe Arg Gly | Asn Val Asn Gly Thr | Phe | | |
| | 290 | | 295 | | 300 |
| Ile Ile His Pro | Asp Ser Gly Asn Leu | Thr Val Ala Arg Ser | Val | | |
| | 305 | | 310 | | 315 |
| Pro Ser Pro Met | Thr Phe Leu Leu Leu | Val Lys Gly Gln Gln | Ala | | |
| | 320 | | 325 | | 330 |
| Asp Leu Ala Arg | Tyr Ser Val Thr Gln | Val Thr Val Glu Ala | Val | | |
| | 335 | | 340 | | 345 |
| Ala Ala Ala Gly | Ser Pro Pro Arg Phe | Pro Gln Arg Leu Tyr | Arg | | |
| | 350 | | 355 | | 360 |
| Gly Thr Val Ala | Arg Gly Ala Gly Ala | Gly Val Val Val Lys | Asp | | |
| | 365 | | 370 | | 375 |
| Ala Ala Ala Pro | Ser Gln Pro Leu Arg | Ile Gln Ala Gln Asp | Pro | | |
| | 380 | | 385 | | 390 |
| Glu Phe Ser Asp | Leu Asn Ser Ala Ile | Thr Tyr Arg Ile Thr | Asn | | |
| | 395 | | 400 | | 405 |
| His Ser His Phe | Arg Met Glu Gly Glu | Val Val Leu Thr Thr | Thr | | |
| | 410 | | 415 | | 420 |
| Thr Leu Ala Gln | Ala Gly Ala Phe Tyr | Ala Glu Val Glu Ala | His | | |
| | 425 | | 430 | | 435 |
| Asn Thr Val Thr | Ser Gly Thr Ala Thr | Thr Val Ile Glu Ile | Gln | | |
| | 440 | | 445 | | 450 |
| Val Ser Glu Gln | Glu Pro Pro Ser Thr | Ala Gln Thr Pro Glu | Ala | | |
| | 455 | | 460 | | 465 |
| Gly Thr Ser Gln | Pro Met Pro Pro Gly | Met Gly Thr Ser Thr | Ser | | |
| | 470 | | 475 | | 480 |
| His Gln Pro Thr | Thr Pro Gly Gly Gly | Thr Ala Gln Thr Pro | Glu | | |
| | 485 | | 490 | | 495 |
| Pro Gly Thr Ser | Gln Pro Met Pro Leu | Ser Lys Ser Thr Pro | Ser | | |
| | 500 | | 505 | | 510 |
| Ser Gly Gly Gly | Pro Ser Glu Asp Lys | Arg Phe Ser Val Val | Asp | | |
| | 515 | | 520 | | 525 |
| Met Ala Ala Leu | Gly Gly Val Leu Gly | Ala Leu Leu Leu Leu | Ala | | |
| | 530 | | 535 | | 540 |
| Leu Leu Gly Leu | Ala Val Leu Val His | Lys His Tyr Gly Pro | Arg | | |
| | 545 | | 550 | | 555 |
| Leu Lys Cys Cys | Cys Gly Lys Ala Pro | Glu Pro Gln Pro Gln | Gly | | |
| | 560 | | 565 | | 570 |
| Phe Asp Asn Gln | Ala Phe Leu Pro Asp | His Lys Ala Asn Trp | Ala | | |
| | 575 | | 580 | | 585 |
| Pro Val Pro Ser | Pro Thr His Asp Pro | Lys Pro Ala Glu Ala | Pro | | |
| | 590 | | 595 | | 600 |
| Met Pro Ala Glu | Pro Ala Pro Pro Gly | Pro Ala Ser Pro Gly | Gly | | |
| | 605 | | 610 | | 615 |
| Ala Pro Glu Pro | Pro Ala Ala Ala Arg | Ala Gly Gly Ser Pro | Thr | | |
| | 620 | | 625 | | 630 |
| Ala Val Arg Ser | Ile Leu Thr Lys Glu | Arg Arg Pro Glu Gly | Gly | | |
| | 635 | | 640 | | 645 |
| Tyr Lys Ala Val | Trp Phe Gly Glu Asp | Ile Gly Thr Glu Ala | Asp | | |
| | 650 | | 655 | | 660 |
| Val Val Val Leu | Asn Ala Pro Thr Leu | Asp Val Asp Gly Ala | Ser | | |
| | 665 | | 670 | | 675 |

Asp Ser Gly Ser Gly Asp Glu Gly Glu Gly Ala Gly Arg Gly Gly
 680 685 690
 Gly Pro Tyr Asp Ala Pro Gly Gly Asp Asp Ser Tyr Ile
 695 700

<210> 15
 <211> 1032
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7099880CD1

<400> 15
 Met Asp Leu Arg Asp Phe Tyr Leu Leu Ala Ala Leu Ile Ala Cys
 1 5 10 15
 Leu Arg Leu Asp Ser Ala Ile Ala Gln Glu Leu Ile Tyr Thr Ile
 20 25 30
 Arg Glu Glu Leu Pro Glu Asn Val Pro Ile Gly Asn Ile Pro Lys
 35 40 45
 Asp Leu Asn Ile Ser His Ile Asn Ala Ala Thr Gly Thr Ser Ala
 50 55 60
 Ser Leu Val Tyr Arg Leu Val Ser Lys Ala Gly Asp Ala Pro Leu
 65 70 75
 Val Lys Val Ser Ser Ser Thr Gly Glu Ile Phe Thr Thr Ser Asn
 80 85 90
 Arg Ile Asp Arg Glu Lys Leu Cys Ala Gly Ala Ser Tyr Ala Glu
 95 100 105
 Glu Asn Glu Cys Phe Phe Glu Leu Glu Val Val Ile Leu Pro Asn
 110 115 120
 Asp Phe Phe Arg Leu Ile Lys Ile Lys Ile Ile Val Lys Asp Thr
 125 130 135
 Asn Asp Asn Ala Pro Met Phe Pro Ser Pro Val Ile Asn Ile Ser
 140 145 150
 Ile Pro Glu Asn Thr Leu Ile Asn Ser Arg Phe Pro Ile Pro Ser
 155 160 165
 Ala Thr Asp Pro Asp Thr Gly Phe Asn Gly Val Gln His Tyr Glu
 170 175 180
 Leu Leu Asn Gly Gln Ser Val Phe Gly Leu Asp Ile Val Glu Thr
 185 190 195
 Pro Glu Gly Glu Lys Trp Pro Gln Leu Ile Val Gln Gln Asn Leu
 200 205 210
 Asp Arg Glu Gln Lys Asp Thr Tyr Val Met Lys Ile Lys Val Glu
 215 220 225
 Asp Gly Gly Thr Pro Gln Lys Ser Ser Thr Ala Ile Leu Gln Val
 230 235 240
 Thr Val Ser Asp Val Asn Asp Asn Arg Pro Val Phe Lys Glu Gly
 245 250 255
 Gln Val Glu Val His Ile Pro Glu Asn Ala Pro Val Gly Thr Ser
 260 265 270
 Val Ile Gln Leu His Ala Thr Asp Ala Asp Ile Gly Ser Asn Ala
 275 280 285
 Glu Ile Arg Tyr Ile Phe Gly Ala Gln Val Ala Pro Ala Thr Lys
 290 295 300
 Arg Leu Phe Ala Leu Asn Asn Thr Thr Gly Leu Ile Thr Val Gln
 305 310 315
 Arg Ser Leu Asp Arg Glu Glu Thr Ala Ile His Lys Val Thr Val
 320 325 330
 Leu Ala Ser Asp Gly Ser Ser Thr Pro Ala Arg Ala Thr Val Thr
 335 340 345
 Ile Asn Val Thr Asp Val Asn Asp Asn Pro Pro Asn Ile Asp Leu
 350 355 360

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Tyr | Ile | Ile | Ser | Pro | Ile | Asn | Gly | Thr | Val | Tyr | Leu | Ser | Glu |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Lys | Asp | Pro | Val | Asn | Thr | Lys | Ile | Ala | Leu | Ile | Thr | Val | Ser | Asp |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Lys | Asp | Thr | Asp | Val | Asn | Gly | Lys | Val | Ile | Cys | Phe | Ile | Glu | Arg |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Glu | Val | Pro | Phe | His | Leu | Lys | Ala | Val | Tyr | Asp | Asn | Gln | Tyr | Leu |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Leu | Glu | Thr | Ser | Ser | Leu | Leu | Asp | Tyr | Glu | Gly | Thr | Lys | Glu | Phe |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Ser | Phe | Lys | Ile | Val | Ala | Ser | Asp | Ser | Gly | Lys | Pro | Ser | Leu | Asn |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Gln | Thr | Ala | Leu | Val | Arg | Val | Lys | Leu | Glu | Asp | Glu | Asn | Asp | Asn |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Pro | Pro | Ile | Phe | Asn | Gln | Pro | Val | Ile | Glu | Leu | Ser | Val | Ser | Glu |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Asn | Asn | Arg | Arg | Gly | Leu | Tyr | Leu | Thr | Thr | Ile | Ser | Ala | Thr | Asp |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Glu | Asp | Ser | Gly | Lys | Asn | Ala | Asp | Ile | Val | Tyr | Gln | Leu | Gly | Pro |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Asn | Ala | Ser | Phe | Phe | Asp | Leu | Asp | Arg | Lys | Thr | Gly | Val | Leu | Thr |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Ala | Ser | Arg | Val | Phe | Asp | Arg | Glu | Glu | Gln | Glu | Arg | Phe | Ile | Phe |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Thr | Val | Thr | Ala | Arg | Asp | Asn | Gly | Thr | Pro | Pro | Leu | Gln | Ser | Gln |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Ala | Ala | Val | Ile | Val | Thr | Val | Leu | Asp | Glu | Asn | Asp | Asn | Ser | Pro |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Lys | Phe | Thr | His | Asn | His | Phe | Gln | Phe | Phe | Val | Ser | Glu | Asn | Leu |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Pro | Lys | Tyr | Ser | Thr | Val | Gly | Val | Ile | Thr | Val | Thr | Asp | Ala | Asp |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Ala | Gly | Glu | Asn | Lys | Ala | Val | Thr | Leu | Ser | Ile | Leu | Asn | Asp | Asn |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Asp | Asn | Phe | Val | Leu | Asp | Pro | Tyr | Ser | Gly | Val | Ile | Lys | Ser | Asn |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Val | Ser | Phe | Asp | Arg | Glu | Gln | Gln | Ser | Ser | Tyr | Thr | Phe | Asp | Val |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Lys | Ala | Thr | Asp | Gly | Gly | Gln | Pro | Pro | Arg | Ser | Ser | Thr | Ala | Lys |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Val | Thr | Ile | Asn | Val | Met | Asp | Val | Asn | Asp | Asn | Ser | Pro | Val | Val |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Ile | Ser | Pro | Pro | Ser | Asn | Thr | Ser | Phe | Lys | Leu | Val | Pro | Leu | Ser |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Ala | Ile | Pro | Gly | Ser | Val | Val | Ala | Glu | Val | Phe | Ala | Val | Asp | Val |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Asp | Thr | Gly | Met | Asn | Ala | Glu | Leu | Lys | Tyr | Thr | Ile | Val | Ser | Gly |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Asn | Asn | Lys | Gly | Leu | Phe | Arg | Ile | Asp | Pro | Val | Thr | Gly | Asn | Ile |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Thr | Leu | Glu | Glu | Lys | Pro | Ala | Pro | Thr | Asp | Val | Gly | Leu | His | Arg |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Leu | Val | Val | Asn | Ile | Ser | Asp | Leu | Gly | Tyr | Pro | Lys | Ser | Leu | His |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Thr | Leu | Val | Leu | Val | Phe | Leu | Tyr | Val | Asn | Asp | Thr | Ala | Gly | Asn |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Ala | Ser | Tyr | Ile | Tyr | Asp | Leu | Ile | Arg | Arg | Thr | Met | Glu | Thr | Pro |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Leu | Asp | Arg | Asn | Ile | Gly | Asp | Ser | Ser | Gln | Pro | Tyr | Gln | Asn | Glu |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Asp | Tyr | Leu | Thr | Ile | Met | Ile | Ala | Ile | Ile | Ala | Gly | Ala | Met | Val |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Val | Ile | Val | Val | Ile | Phe | Val | Thr | Val | Leu | Val | Arg | Cys | Arg | His |

| | | | | | |
|-----------------|------|---------------------|------|---------------------|------|
| Ala Ser Arg Phe | 830 | Lys Ala Ala Gln Arg | 835 | Ser Lys Gln Gly Ala | 840 |
| | 845 | | 850 | | 855 |
| Trp Met Ser Pro | 860 | Asn Gln Glu Asn Lys | 865 | Gln Asn Lys Lys Lys | 870 |
| Arg Lys Lys Arg | 875 | Lys Ser Pro Lys Ser | 880 | Ser Leu Leu Asn Phe | 885 |
| Thr Ile Glu Glu | 890 | Ser Lys Pro Asp Asp | 895 | Ala Val His Glu Pro | 900 |
| Asn Gly Thr Ile | 905 | Ser Leu Pro Ala Glu | 910 | Leu Glu Glu Gln Ser | 915 |
| Gly Arg Phe Asp | 920 | Trp Gly Pro Ala Pro | 925 | Pro Thr Thr Phe Lys | 930 |
| Asn Ser Pro Asp | 935 | Leu Ala Lys His Tyr | 940 | Lys Ser Ala Ser Pro | 945 |
| Pro Ala Phe His | 950 | Leu Lys Pro Asp Thr | 955 | Pro Val Ser Val Lys | 960 |
| His His Val Ile | 965 | Gln Glu Leu Pro Leu | 970 | Asp Asn Thr Phe Val | 975 |
| Gly Cys Asp Thr | 980 | Leu Ser Lys Arg Ser | 985 | Ser Thr Ser Ser Asp | 990 |
| Phe Ser Ala Ser | 995 | Glu Cys Ser Ser Gln | 1000 | Gly Gly Phe Lys Thr | 1005 |
| Gly Pro Leu His | 1010 | Thr Arg Gln Val Asn | 1015 | Glu His Phe Tyr Trp | 1020 |
| Ile Ser Thr Ala | 1025 | Tyr Lys Cys Pro Val | 1030 | Asn Gln Tyr | |

<210> 16
 <211> 687
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 871513CD1

<400> 16

| | | |
|---------------------|---------------------|---------------------|
| Met Ala Leu Asp Gly | Ile Arg Met Pro Asp | Gly Cys Tyr Ala Asp |
| 1 5 | 10 | 15 |
| Gly Thr Trp Glu Leu | Ser Val His Val Thr | Asp Leu Asn Arg Asp |
| 20 | 25 | 30 |
| Val Thr Leu Arg Val | Thr Gly Glu Val His | Ile Gly Gly Val Met |
| 35 | 40 | 45 |
| Leu Lys Leu Val Glu | Lys Leu Asp Val Lys | Lys Asp Trp Ser Asp |
| 50 | 55 | 60 |
| His Ala Leu Trp Trp | Glu Lys Lys Arg Thr | Trp Leu Leu Lys Thr |
| 65 | 70 | 75 |
| His Trp Thr Leu Asp | Lys Tyr Gly Ile Gln | Ala Asp Ala Lys Leu |
| 80 | 85 | 90 |
| Gln Phe Thr Pro Gln | His Lys Leu Leu Arg | Leu Gln Leu Pro Asn |
| 95 | 100 | 105 |
| Met Lys Tyr Val Lys | Val Lys Val Asn Phe | Ser Asp Arg Val Phe |
| 110 | 115 | 120 |
| Lys Ala Val Ser Asp | Ile Cys Lys Thr Phe | Asn Ile Arg His Pro |
| 125 | 130 | 135 |
| Glu Glu Leu Ser Leu | Leu Lys Lys Pro Arg | Asp Pro Thr Lys Lys |
| 140 | 145 | 150 |
| Lys Lys Lys Lys Leu | Asp Asp Gln Ser Glu | Asp Glu Ala Leu Glu |
| 155 | 160 | 165 |
| Leu Glu Gly Pro Leu | Ile Thr Pro Gly Ser | Gly Ser Ile Tyr Ser |
| 170 | 175 | 180 |
| Ser Pro Gly Leu Tyr | Ser Lys Thr Met Thr | Pro Thr Tyr Asp Ala |

| | | | | | |
|-----------------|---------------------|-------------------------|-----|--|-----|
| | 185 | | 190 | | 195 |
| His Asp Gly Ser | Pro Leu Ser Pro Thr | Ser Ala Trp Phe Gly Asp | | | |
| | 200 | | 205 | | 210 |
| Ser Ala Leu Ser | Glu Gly Asn Pro Gly | Ile Leu Ala Val Ser Gln | | | |
| | 215 | | 220 | | 225 |
| Pro Ile Thr Ser | Pro Glu Ile Leu Ala | Lys Met Phe Lys Pro Gln | | | |
| | 230 | | 235 | | 240 |
| Ala Leu Leu Asp | Lys Ala Lys Ile Asn | Gln Gly Trp Leu Asp Ser | | | |
| | 245 | | 250 | | 255 |
| Ser Arg Ser Leu | Met Glu Gln Asp Val | Lys Glu Asn Glu Ala Leu | | | |
| | 260 | | 265 | | 270 |
| Leu Leu Arg Phe | Lys Tyr Tyr Ser Phe | Phe Asp Leu Asn Pro Lys | | | |
| | 275 | | 280 | | 285 |
| Tyr Asp Ala Ile | Arg Ile Asn Gln Leu | Tyr Glu Gln Ala Lys Trp | | | |
| | 290 | | 295 | | 300 |
| Ala Ile Leu Leu | Glu Glu Ile Glu Cys | Thr Glu Glu Glu Met Met | | | |
| | 305 | | 310 | | 315 |
| Met Phe Ala Ala | Leu Gln Tyr His Ile | Asn Lys Leu Ser Ile Met | | | |
| | 320 | | 325 | | 330 |
| Thr Ser Glu Asn | His Leu Asn Asn Ser | Asp Lys Glu Val Asp Glu | | | |
| | 335 | | 340 | | 345 |
| Val Asp Ala Ala | Leu Ser Asp Leu Glu | Ile Thr Leu Glu Gly Gly | | | |
| | 350 | | 355 | | 360 |
| Lys Thr Ser Thr | Ile Leu Gly Asp Ile | Thr Ser Ile Pro Glu Leu | | | |
| | 365 | | 370 | | 375 |
| Ala Asp Tyr Ile | Lys Val Phe Lys Pro | Lys Lys Leu Thr Leu Lys | | | |
| | 380 | | 385 | | 390 |
| Gly Tyr Lys Gln | Tyr Trp Cys Thr Phe | Lys Asp Thr Ser Ile Ser | | | |
| | 395 | | 400 | | 405 |
| Cys Tyr Lys Ser | Lys Glu Glu Ser Ser | Gly Thr Pro Ala His Gln | | | |
| | 410 | | 415 | | 420 |
| Met Asn Leu Arg | Gly Cys Glu Val Thr | Pro Asp Val Asn Ile Ser | | | |
| | 425 | | 430 | | 435 |
| Gly Gln Lys Phe | Asn Ile Lys Leu Leu | Ile Pro Val Ala Glu Gly | | | |
| | 440 | | 445 | | 450 |
| Met Asn Glu Ile | Trp Leu Arg Cys Asp | Asn Glu Lys Gln Tyr Ala | | | |
| | 455 | | 460 | | 465 |
| His Trp Met Ala | Ala Cys Arg Leu Ala | Ser Lys Gly Lys Thr Met | | | |
| | 470 | | 475 | | 480 |
| Ala Asp Ser Ser | Tyr Asn Leu Glu Val | Gln Asn Ile Leu Ser Phe | | | |
| | 485 | | 490 | | 495 |
| Leu Lys Met Gln | His Leu Asn Pro Asp | Pro Gln Leu Ile Pro Glu | | | |
| | 500 | | 505 | | 510 |
| Gln Ile Thr Thr | Asp Ile Thr Pro Glu | Cys Leu Val Ser Pro Arg | | | |
| | 515 | | 520 | | 525 |
| Tyr Leu Lys Lys | Tyr Lys Asn Lys Gln | Pro Gly Tyr Ile Arg Asp | | | |
| | 530 | | 535 | | 540 |
| Leu Ile Thr Ala | Arg Ile Leu Glu Ala | His Gln Asn Val Ala Gln | | | |
| | 545 | | 550 | | 555 |
| Met Ser Leu Ile | Glu Ala Lys Met Arg | Phe Ile Gln Ala Trp Gln | | | |
| | 560 | | 565 | | 570 |
| Ser Leu Pro Glu | Phe Gly Ile Thr His | Phe Ile Ala Arg Phe Gln | | | |
| | 575 | | 580 | | 585 |
| Gly Gly Lys Lys | Glu Glu Leu Ile Gly | Ile Ala Tyr Asn Arg Leu | | | |
| | 590 | | 595 | | 600 |
| Ile Arg Met Asp | Ala Ser Thr Gly Asp | Ala Ile Lys Thr Trp Arg | | | |
| | 605 | | 610 | | 615 |
| Phe Ser Asn Met | Lys Gln Trp Asn Val | Asn Trp Glu Ile Lys Met | | | |
| | 620 | | 625 | | 630 |
| Val Thr Val Glu | Phe Ala Asp Glu Val | Arg Leu Ser Phe Ile Cys | | | |
| | 635 | | 640 | | 645 |
| Thr Glu Val Asp | Cys Lys Val Val His | Glu Phe Ile Gly Gly Tyr | | | |
| | 650 | | 655 | | 660 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Phe | Leu | Ser | Thr | Arg | Ala | Lys | Asp | Gln | Asn | Glu | Ser | Leu | Asp |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Glu | Glu | Met | Phe | Tyr | Lys | Leu | Thr | Ser | Gly | Trp | Val | | | |
| | | | | 680 | | | | | 685 | | | | | |

<210> 17
 <211> 1805
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 8057640CD1

<400> 17

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Gly | Pro | Arg | Pro | Ser | Pro | Trp | Ala | Arg | Leu | Leu | Leu | Ala |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Ala | Leu | Ile | Ser | Val | Ser | Leu | Ser | Gly | Thr | Leu | Ala | Asn | Arg | Cys |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Lys | Lys | Ala | Pro | Val | Lys | Ser | Cys | Thr | Glu | Cys | Val | Arg | Val | Asp |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Lys | Asp | Cys | Ala | Tyr | Cys | Thr | Asp | Glu | Met | Phe | Arg | Asp | Arg | Arg |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Cys | Asn | Thr | Gln | Ala | Glu | Leu | Leu | Ala | Ala | Gly | Cys | Gln | Arg | Glu |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Ser | Ile | Val | Val | Met | Glu | Ser | Ser | Phe | Gln | Ile | Thr | Glu | Glu | Thr |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Gln | Ile | Asp | Thr | Thr | Leu | Arg | Arg | Ser | Gln | Met | Ser | Pro | Gln | Gly |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Leu | Arg | Val | Arg | Leu | Arg | Pro | Gly | Glu | Glu | Arg | His | Phe | Glu | Leu |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Glu | Val | Phe | Glu | Pro | Leu | Glu | Ser | Pro | Val | Asp | Leu | Tyr | Ile | Leu |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Met | Asp | Phe | Ser | Asn | Ser | Met | Ser | Asp | Asp | Leu | Asp | Asn | Leu | Lys |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Lys | Met | Gly | Gln | Asn | Leu | Ala | Arg | Val | Leu | Ser | Gln | Leu | Thr | Ser |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Asp | Tyr | Thr | Ile | Gly | Phe | Gly | Lys | Phe | Val | Asp | Lys | Val | Ser | Val |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Pro | Gln | Thr | Asp | Met | Arg | Pro | Glu | Lys | Leu | Lys | Glu | Pro | Trp | Pro |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Asn | Ser | Asp | Pro | Pro | Phe | Ser | Phe | Lys | Asn | Val | Ile | Ser | Leu | Thr |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Glu | Asp | Val | Asp | Glu | Phe | Arg | Asn | Lys | Leu | Gln | Gly | Glu | Arg | Ile |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Ser | Gly | Asn | Leu | Asp | Ala | Pro | Glu | Gly | Gly | Phe | Asp | Ala | Ile | Leu |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Gln | Thr | Ala | Val | Cys | Thr | Arg | Asp | Ile | Gly | Trp | Arg | Pro | Asp | Ser |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Thr | His | Leu | Leu | Val | Phe | Ser | Thr | Glu | Ser | Ala | Phe | His | Tyr | Glu |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Ala | Asp | Gly | Ala | Asn | Val | Leu | Ala | Gly | Ile | Met | Ser | Arg | Asn | Asp |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Glu | Arg | Cys | His | Leu | Asp | Thr | Thr | Gly | Thr | Tyr | Thr | Gln | Tyr | Arg |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Thr | Gln | Asp | Tyr | Pro | Ser | Val | Pro | Thr | Leu | Val | Arg | Leu | Leu | Ala |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Lys | His | Asn | Ile | Ile | Pro | Ile | Phe | Ala | Val | Thr | Asn | Tyr | Ser | Tyr |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Ser | Tyr | Tyr | Glu | Lys | Leu | His | Thr | Tyr | Phe | Pro | Val | Ser | Ser | Leu |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Gly | Val | Leu | Gln | Glu | Asp | Ser | Ser | Asn | Ile | Val | Glu | Leu | Leu | Glu |
| | | | | 350 | | | | | 355 | | | | | 360 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Glu | Ala | Phe | Asn | Arg | Ile | Arg | Ser | Asn | Leu | Asp | Ile | Arg | Ala | Leu |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Asp | Ser | Pro | Arg | Gly | Leu | Arg | Thr | Glu | Val | Thr | Ser | Lys | Met | Phe |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Gln | Lys | Thr | Arg | Thr | Gly | Ser | Phe | His | Ile | Arg | Arg | Gly | Glu | Val |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Gly | Ile | Tyr | Gln | Val | Gln | Leu | Arg | Ala | Leu | Glu | His | Val | Asp | Gly |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Thr | His | Val | Cys | Gln | Leu | Pro | Glu | Asp | Gln | Lys | Gly | Asn | Ile | His |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Leu | Lys | Pro | Ser | Phe | Ser | Asp | Gly | Leu | Lys | Met | Asp | Ala | Gly | Ile |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Ile | Cys | Asp | Val | Cys | Thr | Cys | Glu | Leu | Gln | Lys | Glu | Val | Arg | Ser |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Ala | Arg | Cys | Ser | Phe | Asn | Gly | Asp | Phe | Val | Cys | Gly | Gln | Cys | Val |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Cys | Ser | Glu | Gly | Trp | Ser | Gly | Gln | Thr | Cys | Asn | Cys | Ser | Thr | Gly |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Ser | Leu | Ser | Asp | Ile | Gln | Pro | Cys | Leu | Arg | Glu | Gly | Glu | Asp | Lys |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Pro | Cys | Ser | Gly | Arg | Gly | Glu | Cys | Gln | Cys | Gly | His | Cys | Val | Cys |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Tyr | Gly | Glu | Gly | Arg | Tyr | Glu | Gly | Gln | Phe | Cys | Glu | Tyr | Asp | Asn |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Phe | Gln | Cys | Pro | Arg | Thr | Ser | Gly | Phe | Leu | Cys | Asn | Asp | Arg | Gly |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Arg | Cys | Ser | Met | Gly | Gln | Cys | Val | Cys | Glu | Pro | Gly | Trp | Thr | Gly |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Pro | Ser | Cys | Asp | Cys | Pro | Leu | Ser | Asn | Ala | Thr | Cys | Ile | Asp | Ser |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Asn | Gly | Gly | Ile | Cys | Asn | Gly | Arg | Gly | His | Cys | Glu | Cys | Gly | Arg |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Cys | His | Cys | His | Gln | Gln | Ser | Leu | Tyr | Thr | Asp | Thr | Ile | Cys | Glu |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Ile | Asn | Tyr | Ser | Ala | Ile | His | Pro | Gly | Leu | Cys | Glu | Asp | Leu | Arg |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Ser | Cys | Val | Gln | Cys | Gln | Ala | Trp | Gly | Thr | Gly | Glu | Lys | Lys | Gly |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Arg | Thr | Cys | Glu | Glu | Cys | Asn | Phe | Lys | Val | Lys | Met | Val | Asp | Glu |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Leu | Lys | Arg | Ala | Glu | Glu | Val | Val | Val | Arg | Cys | Ser | Phe | Arg | Asp |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Glu | Asp | Asp | Asp | Cys | Thr | Tyr | Ser | Tyr | Thr | Met | Glu | Gly | Asp | Gly |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Ala | Pro | Gly | Pro | Asn | Ser | Thr | Val | Leu | Val | His | Lys | Lys | Lys | Asp |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Cys | Pro | Pro | Gly | Ser | Phe | Trp | Trp | Leu | Ile | Pro | Leu | Leu | Leu | Leu |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Leu | Leu | Pro | Leu | Leu | Ala | Leu | Leu | Leu | Leu | Leu | Cys | Trp | Lys | Tyr |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Cys | Ala | Cys | Cys | Lys | Ala | Cys | Leu | Ala | Leu | Leu | Pro | Cys | Cys | Asn |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Arg | Gly | His | Met | Val | Gly | Phe | Lys | Glu | Asp | His | Tyr | Met | Leu | Arg |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Glu | Asn | Leu | Met | Ala | Ser | Asp | His | Leu | Asp | Thr | Pro | Met | Leu | Arg |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Ser | Gly | Asn | Leu | Lys | Gly | Arg | Asp | Val | Val | Arg | Trp | Lys | Val | Thr |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Asn | Asn | Met | Gln | Arg | Pro | Gly | Phe | Ala | Thr | His | Ala | Ala | Ser | Ile |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Asn | Pro | Thr | Glu | Leu | Val | Pro | Tyr | Gly | Leu | Ser | Leu | Arg | Leu | Ala |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Arg | Leu | Cys | Thr | Glu | Asn | Leu | Leu | Lys | Pro | Asp | Thr | Arg | Glu | Cys |

| | | | | | |
|-----------------|------|---------------------|------|---------------------|------|
| Ala Gln Leu Arg | 830 | Gln Glu Val Glu Glu | 835 | Leu Asn Glu Val | 840 |
| | 845 | | 850 | | 855 |
| Arg Gln Ile Ser | 860 | Gly Val His Lys Leu | 865 | Gln Gln Thr Lys Phe | 870 |
| | 875 | Ala Gly Lys Lys Gln | 880 | Asp His Thr Ile Val | 885 |
| Thr Val Leu Met | 890 | Ala Pro Arg Ser Ala | 895 | Pro Ala Leu Leu | 900 |
| | 905 | | 910 | | 915 |
| Val Ala Pro Gly | 920 | Tyr Tyr Thr Leu Thr | 925 | Ala Asp Gln Asp Ala | 930 |
| | 935 | | 940 | | 945 |
| Gly Met Val Glu | 950 | Phe Gln Glu Gly Val | 955 | Leu Val Asp Val | 960 |
| | 965 | | 970 | | 975 |
| Val Pro Leu Phe | 980 | Ile Arg Pro Glu Asp | 985 | Asp Asp Glu Lys Gln | 990 |
| | 995 | | 1000 | | 1005 |
| Leu Val Glu Ala | 1010 | Ile Asp Val Pro Ala | 1015 | Gly Thr Ala Thr Leu | 1020 |
| | 1025 | | 1030 | | 1035 |
| Arg Arg Leu Val | 1040 | Asn Ile Thr Ile Ile | 1045 | Lys Glu Gln Ala Arg | 1050 |
| | 1055 | | 1060 | | 1065 |
| Val Val Ser Phe | 1070 | Glu Gln Pro Glu Phe | 1075 | Ser Val Ser Arg Gly | 1080 |
| | 1085 | | 1090 | | 1095 |
| Gln Val Ala Arg | 1100 | Ile Pro Val Ile Arg | 1105 | Arg Val Leu Asp Gly | 1110 |
| | 1115 | | 1120 | | 1125 |
| Lys Ser Gln Val | 1130 | Ser Tyr Arg Thr Gln | 1135 | Asp Gly Thr Ala Gln | 1140 |
| | 1145 | | 1150 | | 1155 |
| Asn Arg Asp Tyr | 1160 | Ile Pro Val Glu Gly | 1165 | Leu Leu Phe Gln Pro | 1170 |
| | 1175 | | 1180 | | 1185 |
| Gly Glu Ala Trp | 1190 | Lys Glu Leu Gln Val | 1195 | Lys Leu Leu Glu Leu | 1200 |
| | 1205 | | 1210 | | 1215 |
| Glu Val Asp Ser | 1220 | Leu Leu Arg Gly Arg | 1225 | Gln Val Arg Arg Phe | 1230 |
| | 1235 | | 1240 | | 1245 |
| Val Gln Leu Ser | 1250 | Asn Pro Lys Phe Gly | 1255 | Ala His Leu Gly Gln | 1260 |
| | 1265 | | 1270 | | 1275 |
| His Ser Thr Thr | 1280 | Ile Ile Ile Arg Asp | 1285 | Pro Asp Glu Leu Asp | 1290 |
| | 1295 | | 1300 | | 1305 |
| Ser Phe Thr Ser | | Gln Met Leu Ser Ser | | Gln Pro Pro Pro His | |
| | | | | | |
| Asp Leu Gly Ala | | Pro Gln Asn Pro Asn | | Ala Lys Ala Ala Gly | |
| | | | | | |
| Arg Lys Ile His | | Phe Asn Trp Leu Pro | | Pro Ser Gly Lys Pro | |
| | | | | | |
| Gly Tyr Arg Val | | Lys Tyr Trp Ile Gln | | Gly Asp Ser Glu Ser | |
| | | | | | |
| Ala His Leu Leu | | Asp Ser Lys Val Pro | | Ser Val Glu Leu Thr | |
| | | | | | |
| Leu Tyr Pro Tyr | | Cys Asp Tyr Glu Met | | Lys Val Cys Ala Tyr | |
| | | | | | |
| Ala Gln Gly Glu | | Gly Pro Tyr Ser Ser | | Leu Val Ser Cys Arg | |
| | | | | | |
| His Gln Glu Val | | Pro Ser Glu Pro Gly | | Arg Leu Ala Phe Asn | |
| | | | | | |
| Val Ser Ser Thr | | Val Thr Gln Leu Ser | | Trp Ala Glu Pro Ala | |
| | | | | | |
| Thr Asn Gly Glu | | Ile Thr Ala Tyr Glu | | Val Cys Tyr Gly Leu | |
| | | | | | |
| Asn Asp Asp Asn | | Arg Pro Ile Gly Pro | | Met Lys Lys Val Leu | |
| | | | | | |
| Asp Asn Pro Lys | | Asn Arg Met Leu Leu | | Ile Glu Asn Leu Arg | |
| | | | | | |
| Ser Gln Pro Tyr | | Arg Tyr Thr Val Lys | | Ala Arg Asn Gly Ala | |
| | | | | | |

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|
| Trp | Gly | Pro | Glu | Arg | Glu | Ala | Ile | Ile | Asn | Leu | Ala | Thr | Gln | Pro | 1310 | 1315 | 1320 |
| Lys | Arg | Pro | Met | Ser | Ile | Pro | Ile | Ile | Pro | Asp | Ile | Pro | Ile | Val | 1325 | 1330 | 1335 |
| Asp | Ala | Gln | Ser | Gly | Glu | Asp | Tyr | Asp | Ser | Phe | Leu | Met | Tyr | Ser | 1340 | 1345 | 1350 |
| Asp | Asp | Val | Leu | Arg | Ser | Pro | Ser | Gly | Ser | Gln | Arg | Pro | Ser | Val | 1355 | 1360 | 1365 |
| Ser | Asp | Asp | Thr | Glu | His | Leu | Val | Asn | Gly | Arg | Met | Asp | Phe | Ala | 1370 | 1375 | 1380 |
| Phe | Pro | Gly | Ser | Thr | Asn | Ser | Leu | His | Arg | Met | Thr | Thr | Thr | Ser | 1385 | 1390 | 1395 |
| Ala | Ala | Ala | Tyr | Gly | Thr | His | Leu | Ser | Pro | His | Val | Pro | His | Arg | 1400 | 1405 | 1410 |
| Val | Leu | Ser | Thr | Ser | Ser | Thr | Leu | Thr | Arg | Asp | Tyr | Asn | Ser | Leu | 1415 | 1420 | 1425 |
| Thr | Arg | Ser | Glu | His | Ser | His | Ser | Thr | Thr | Leu | Pro | Arg | Asp | Tyr | 1430 | 1435 | 1440 |
| Ser | Thr | Leu | Thr | Ser | Val | Ser | Ser | His | Gly | Leu | Pro | Pro | Ile | Trp | 1445 | 1450 | 1455 |
| Glu | His | Gly | Arg | Ser | Arg | Leu | Pro | Leu | Ser | Trp | Ala | Leu | Gly | Ser | 1460 | 1465 | 1470 |
| Arg | Ser | Arg | Ala | Gln | Met | Lys | Gly | Phe | Pro | Pro | Ser | Arg | Gly | Pro | 1475 | 1480 | 1485 |
| Arg | Asp | Ser | Ile | Ile | Leu | Ala | Gly | Arg | Pro | Ala | Ala | Pro | Ser | Trp | 1490 | 1495 | 1500 |
| Gly | Pro | Asp | Ser | Arg | Leu | Thr | Ala | Gly | Val | Pro | Asp | Thr | Pro | Thr | 1505 | 1510 | 1515 |
| Arg | Leu | Val | Phe | Ser | Ala | Leu | Gly | Pro | Thr | Ser | Leu | Arg | Val | Ser | 1520 | 1525 | 1530 |
| Trp | Gln | Glu | Pro | Arg | Cys | Glu | Arg | Pro | Leu | Gln | Gly | Tyr | Ser | Val | 1535 | 1540 | 1545 |
| Glu | Tyr | Gln | Leu | Leu | Asn | Gly | Gly | Glu | Leu | His | Arg | Leu | Asn | Ile | 1550 | 1555 | 1560 |
| Pro | Asn | Pro | Ala | Gln | Thr | Ser | Val | Val | Val | Glu | Asp | Leu | Leu | Pro | 1565 | 1570 | 1575 |
| Asn | His | Ser | Tyr | Val | Phe | Arg | Val | Arg | Ala | Gln | Ser | Gln | Glu | Gly | 1580 | 1585 | 1590 |
| Trp | Gly | Arg | Glu | Arg | Glu | Gly | Val | Ile | Thr | Ile | Glu | Ser | Gln | Val | 1595 | 1600 | 1605 |
| His | Pro | Gln | Ser | Pro | Leu | Cys | Pro | Leu | Pro | Gly | Ser | Ala | Phe | Thr | 1610 | 1615 | 1620 |
| Leu | Ser | Thr | Pro | Ser | Ala | Pro | Gly | Pro | Leu | Val | Phe | Thr | Ala | Leu | 1625 | 1630 | 1635 |
| Ser | Pro | Asp | Ser | Leu | Gln | Leu | Ser | Trp | Glu | Arg | Pro | Arg | Arg | Pro | 1640 | 1645 | 1650 |
| Asn | Gly | Asp | Ile | Val | Gly | Tyr | Leu | Val | Thr | Cys | Glu | Met | Ala | Gln | 1655 | 1660 | 1665 |
| Gly | Gly | Gly | Pro | Ala | Thr | Ala | Phe | Arg | Val | Asp | Gly | Asp | Ser | Pro | 1670 | 1675 | 1680 |
| Glu | Ser | Arg | Leu | Thr | Val | Pro | Gly | Leu | Ser | Glu | Asn | Val | Pro | Tyr | 1685 | 1690 | 1695 |
| Lys | Phe | Lys | Val | Gln | Ala | Arg | Thr | Thr | Glu | Gly | Phe | Gly | Pro | Glu | 1700 | 1705 | 1710 |
| Arg | Glu | Gly | Ile | Ile | Thr | Ile | Glu | Ser | Gln | Asp | Gly | Gly | Pro | Phe | 1715 | 1720 | 1725 |
| Pro | Gln | Leu | Gly | Ser | Arg | Ala | Gly | Leu | Phe | Gln | His | Pro | Leu | Gln | 1730 | 1735 | 1740 |
| Ser | Glu | Tyr | Ser | Ser | Ile | Thr | Thr | Thr | His | Thr | Ser | Ala | Thr | Glu | 1745 | 1750 | 1755 |
| Pro | Phe | Leu | Val | Asp | Gly | Leu | Thr | Leu | Gly | Ala | Gln | His | Leu | Glu | 1760 | 1765 | 1770 |
| Thr | Gly | Gly | Ser | Leu | Thr | Arg | His | Val | Thr | Gln | Glu | Phe | Val | Ser | | | |

| | | | | | |
|---------------------|---------------------|---------------------|------|------|------|
| | 1775 | | 1780 | | 1785 |
| Arg Thr Leu Thr Thr | Ser Gly Thr Leu Ser | Thr His Met Asp Gln | | | |
| | 1790 | 1795 | | 1800 | |
| Gln Phe Phe Gln Thr | | | | | |
| | 1805 | | | | |

<210> 18

<211> 372

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7505913CD1

<400> 18

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Asp | Gly | Val | Tyr | Glu | Pro | Pro | Asp | Leu | Thr | Pro | Glu | Glu |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Arg | Met | Glu | Leu | Glu | Asn | Ile | Arg | Arg | Arg | Lys | Gln | Glu | Leu | Leu |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Val | Glu | Ile | Gln | Arg | Leu | Arg | Glu | Glu | Leu | Ser | Glu | Ala | Met | Ser |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Glu | Val | Glu | Gly | Leu | Glu | Ala | Asn | Glu | Gly | Ser | Lys | Thr | Leu | Gln |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Arg | Asn | Arg | Lys | Met | Ala | Met | Gly | Arg | Lys | Lys | Phe | Asn | Met | Asp |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Pro | Lys | Lys | Gly | Ile | Gln | Phe | Leu | Val | Glu | Asn | Glu | Leu | Leu | Gln |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Asn | Thr | Pro | Glu | Glu | Ile | Ala | Arg | Phe | Leu | Tyr | Lys | Gly | Glu | Gly |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Leu | Asn | Lys | Thr | Ala | Ile | Gly | Asp | Tyr | Leu | Gly | Glu | Arg | Glu | Glu |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Leu | Asn | Leu | Ala | Val | Leu | His | Ala | Phe | Val | Asp | Leu | His | Glu | Phe |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Thr | Asp | Leu | Asn | Leu | Val | Gln | Ala | Leu | Arg | Gln | Phe | Leu | Trp | Ser |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Phe | Arg | Leu | Pro | Gly | Glu | Ala | Gln | Lys | Ile | Asp | Arg | Met | Met | Glu |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Ala | Phe | Ala | Gln | Arg | Tyr | Cys | Leu | Cys | Asn | Pro | Gly | Val | Phe | Gln |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Ser | Thr | Asp | Thr | Cys | Tyr | Val | Leu | Ser | Phe | Ala | Val | Ile | Met | Leu |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Asn | Thr | Ser | Leu | His | Asn | Pro | Asn | Val | Arg | Asp | Lys | Pro | Gly | Leu |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Glu | Arg | Phe | Val | Ala | Met | Asn | Arg | Gly | Ile | Asn | Glu | Gly | Gly | Asp |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Leu | Pro | Glu | Glu | Leu | Leu | Arg | Asn | Leu | Tyr | Asp | Ser | Ile | Arg | Asn |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Glu | Pro | Phe | Lys | Ile | Pro | Glu | Asp | Asp | Gly | Asn | Asp | Leu | Thr | His |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Thr | Phe | Phe | Asn | Pro | Asp | Arg | Glu | Gly | Trp | Leu | Leu | Lys | Leu | Gly |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Gly | Gly | Arg | Val | Lys | Thr | Trp | Lys | Arg | Arg | Trp | Phe | Ile | Leu | Thr |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asp | Asn | Cys | Leu | Tyr | Tyr | Phe | Glu | Tyr | Thr | Thr | Asp | Lys | Glu | Pro |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Arg | Gly | Ile | Ile | Pro | Leu | Glu | Asn | Leu | Ser | Ile | Arg | Glu | Val | Asp |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Asp | Pro | Arg | Lys | Pro | Asn | Cys | Phe | Glu | Leu | Tyr | Ile | Pro | Asn | Asn |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Lys | Gly | Gln | Glu | Glu | Lys | Asp | Glu | Trp | Ile | Lys | Ser | Ile | Gln | Ala |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Ala | Val | Ser | Val | Asp | Pro | Phe | Tyr | Glu | Met | Leu | Ala | Ala | Arg | Lys |

| | | | | | |
|---------------------|-----------------|-------------|-----|--|-----|
| | 350 | | 355 | | 360 |
| Lys Arg Ile Ser Val | Lys Lys Lys Gln | Glu Gln Pro | | | |
| | 365 | 370 | | | |

<210> 19

<211> 1088

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510292CD1

<400> 19

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Asn | Trp | Thr | Gly | Arg | Pro | Trp | Leu | Tyr | Leu | Leu | Leu | Leu |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Leu | Ser | Leu | Pro | Gln | Leu | Cys | Leu | Asp | Gln | Glu | Val | Leu | Ser | Gly |
| | | | | 20 | | | | | 25 | | | | | 30 |
| His | Ser | Leu | Gln | Thr | Pro | Thr | Glu | Glu | Gly | Gln | Gly | Pro | Glu | Gly |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Val | Trp | Gly | Pro | Trp | Val | Gln | Trp | Ala | Ser | Cys | Ser | Gln | Pro | Cys |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Gly | Val | Gly | Val | Gln | Arg | Arg | Ser | Arg | Thr | Cys | Gln | Leu | Pro | Thr |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Val | Gln | Leu | His | Pro | Ser | Leu | Pro | Leu | Pro | Pro | Arg | Pro | Pro | Arg |
| | | | | 80 | | | | | 85 | | | | | 90 |
| His | Pro | Glu | Ala | Leu | Leu | Pro | Arg | Gly | Gln | Gly | Pro | Arg | Pro | Gln |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Thr | Ser | Pro | Glu | Thr | Leu | Pro | Leu | Tyr | Arg | Thr | Gln | Ser | Arg | Gly |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Arg | Gly | Gly | Pro | Leu | Arg | Gly | Pro | Ala | Ser | His | Leu | Gly | Arg | Glu |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Glu | Thr | Gln | Glu | Ile | Arg | Ala | Ala | Arg | Arg | Ser | Arg | Leu | Arg | Asp |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Pro | Ile | Lys | Pro | Gly | Met | Phe | Gly | Tyr | Gly | Arg | Val | Pro | Phe | Ala |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Leu | Pro | Leu | His | Arg | Asn | Arg | Arg | His | Pro | Arg | Ser | Pro | Pro | Arg |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Ser | Glu | Leu | Ser | Leu | Ile | Ser | Ser | Arg | Gly | Glu | Glu | Pro | Ile | Pro |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Ser | Pro | Thr | Pro | Arg | Ala | Glu | Pro | Phe | Ser | Ala | Asn | Gly | Ser | Pro |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Gln | Thr | Glu | Leu | Pro | Pro | Thr | Glu | Leu | Ser | Val | His | Thr | Pro | Ser |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Pro | Gln | Ala | Glu | Pro | Leu | Ser | Pro | Glu | Thr | Ala | Gln | Thr | Glu | Val |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Ala | Pro | Arg | Thr | Arg | Pro | Ala | Pro | Leu | Arg | His | His | Pro | Arg | Ala |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Gln | Ala | Ser | Gly | Thr | Glu | Pro | Pro | Ser | Pro | Thr | His | Ser | Leu | Gly |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Glu | Gly | Gly | Phe | Phe | Arg | Ala | Ser | Pro | Gln | Pro | Arg | Arg | Pro | Ser |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Ser | Gln | Gly | Trp | Ala | Ser | Pro | Gln | Val | Ala | Gly | Arg | Arg | Pro | Asp |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Pro | Phe | Pro | Ser | Val | Pro | Arg | Gly | Arg | Gly | Gln | Gln | Gly | Gln | Gly |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Pro | Trp | Gly | Thr | Gly | Gly | Thr | Pro | His | Gly | Pro | Arg | Leu | Glu | Pro |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Asp | Pro | Gln | His | Pro | Gly | Ala | Trp | Leu | Pro | Leu | Leu | Ser | Asn | Gly |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Pro | His | Ala | Ser | Ser | Leu | Trp | Ser | Leu | Phe | Ala | Pro | Ser | Ser | Pro |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Ile | Pro | Arg | Cys | Ser | Gly | Glu | Ser | Glu | Gln | Leu | Arg | Ala | Cys | Ser |

| | | | | | |
|-----------------|---------------------|---------------------|-----|--|-----|
| | 365 | | 370 | | 375 |
| Gln Ala Pro Cys | Pro Pro Glu Gln Pro | Asp Pro Arg Ala Leu | Gln | | |
| | 380 | | 385 | | 390 |
| Cys Ala Ala Phe | Asn Ser Gln Glu Phe | Met Gly Gln Leu Tyr | Gln | | |
| | 395 | | 400 | | 405 |
| Trp Glu Pro Phe | Thr Glu Val Gln Gly | Ser Gln Arg Cys Glu | Leu | | |
| | 410 | | 415 | | 420 |
| Asn Cys Arg Pro | Arg Gly Phe Arg Phe | Tyr Val Arg His Thr | Glu | | |
| | 425 | | 430 | | 435 |
| Lys Val Gln Asp | Gly Thr Leu Cys Gln | Pro Gly Ala Pro Asp | Ile | | |
| | 440 | | 445 | | 450 |
| Cys Val Ala Gly | Arg Cys Leu Ser Pro | Gly Cys Asp Gly Ile | Leu | | |
| | 455 | | 460 | | 465 |
| Gly Ser Gly Arg | Arg Pro Asp Gly Cys | Gly Val Cys Gly Gly | Asp | | |
| | 470 | | 475 | | 480 |
| Asp Ser Thr Cys | Arg Leu Val Ser Gly | Asn Leu Thr Asp Arg | Gly | | |
| | 485 | | 490 | | 495 |
| Gly Pro Leu Gly | Tyr Gln Lys Ile Leu | Trp Ile Pro Ala Gly | Ala | | |
| | 500 | | 505 | | 510 |
| Leu Arg Leu Gln | Ile Ala Gln Leu Arg | Pro Ser Ser Asn Tyr | Leu | | |
| | 515 | | 520 | | 525 |
| Ala Leu Arg Gly | Pro Gly Gly Arg Ser | Ile Ile Asn Gly Asn | Trp | | |
| | 530 | | 535 | | 540 |
| Ala Val Asp Pro | Pro Gly Ser Tyr Arg | Ala Gly Gly Thr Val | Phe | | |
| | 545 | | 550 | | 555 |
| Arg Tyr Asn Arg | Pro Pro Arg Glu Glu | Gly Lys Gly Glu Ser | Leu | | |
| | 560 | | 565 | | 570 |
| Ser Ala Glu Gly | Pro Thr Thr Gln Pro | Val Asp Val Tyr Met | Ile | | |
| | 575 | | 580 | | 585 |
| Phe Gln Glu Glu | Asn Pro Gly Val Phe | Tyr Gln Tyr Val Ile | Ser | | |
| | 590 | | 595 | | 600 |
| Ser Pro Pro Pro | Ile Leu Glu Asn Pro | Thr Pro Glu Pro Pro | Val | | |
| | 605 | | 610 | | 615 |
| Pro Gln Leu Gln | Pro Glu Ile Leu Arg | Val Glu Pro Pro Leu | Ala | | |
| | 620 | | 625 | | 630 |
| Pro Ala Pro Arg | Pro Ala Arg Thr Pro | Gly Thr Leu Gln Arg | Gln | | |
| | 635 | | 640 | | 645 |
| Val Arg Ile Pro | Gln Met Pro Ala Pro | Pro His Pro Arg Thr | Pro | | |
| | 650 | | 655 | | 660 |
| Leu Gly Ser Pro | Ala Ala Tyr Trp Lys | Arg Val Gly His Ser | Ala | | |
| | 665 | | 670 | | 675 |
| Cys Ser Ala Ser | Cys Gly Lys Gly Val | Trp Arg Pro Ile Phe | Leu | | |
| | 680 | | 685 | | 690 |
| Cys Ile Ser Arg | Glu Ser Gly Glu Glu | Leu Asp Glu Arg Ser | Cys | | |
| | 695 | | 700 | | 705 |
| Ala Ala Gly Ala | Arg Pro Pro Ala Ser | Pro Glu Pro Cys His | Gly | | |
| | 710 | | 715 | | 720 |
| Thr Pro Cys Pro | Pro Tyr Trp Glu Ala | Gly Glu Trp Thr Ser | Cys | | |
| | 725 | | 730 | | 735 |
| Ser Arg Ser Cys | Gly Pro Gly Thr Gln | His Arg Gln Leu Gln | Cys | | |
| | 740 | | 745 | | 750 |
| Arg Gln Glu Phe | Gly Gly Gly Gly Ser | Ser Val Pro Pro Glu | Arg | | |
| | 755 | | 760 | | 765 |
| Cys Gly His Leu | Pro Arg Pro Asn Ile | Thr Gln Ser Cys Gln | Leu | | |
| | 770 | | 775 | | 780 |
| Arg Leu Cys Gly | His Trp Glu Val Gly | Ser Pro Trp Ser Gln | Val | | |
| | 785 | | 790 | | 795 |
| Trp Glu Ala Gln | Leu Pro Gly Phe Pro | Ser Pro Pro Gln Cys | Ser | | |
| | 800 | | 805 | | 810 |
| Val Arg Cys Gly | Arg Gly Gln Arg Ser | Arg Gln Val Arg Cys | Val | | |
| | 815 | | 820 | | 825 |
| Gly Asn Asn Gly | Asp Glu Val Ser Glu | Gln Glu Cys Ala Ser | Gly | | |
| | 830 | | 835 | | 840 |

```

Pro Pro Gln Pro Pro Ser Arg Glu Ala Cys Asp Met Gly Pro Cys
      845      850      855
Thr Thr Ala Trp Phe His Ser Asp Trp Ser Ser Lys Cys Ser Ala
      860      865      870
Glu Cys Gly Thr Gly Ile Gln Arg Arg Ser Val Val Cys Leu Gly
      875      880      885
Ser Gly Ala Ala Leu Gly Pro Gly Gln Gly Glu Ala Gly Ala Gly
      890      895      900
Thr Gly Gln Ser Cys Pro Thr Gly Ser Arg Pro Pro Asp Met Arg
      905      910      915
Ala Cys Ser Leu Gly Pro Cys Glu Arg Thr Trp Arg Trp Tyr Thr
      920      925      930
Gly Pro Trp Gly Glu Cys Ser Ser Glu Cys Gly Ser Gly Thr Gln
      935      940      945
Arg Arg Asp Ile Cys Val Ser Lys Leu Gly Thr Glu Phe Asn
      950      955      960
Val Thr Ser Pro Ser Asn Cys Ser His Leu Pro Arg Pro Pro Ala
      965      970      975
Leu Gln Pro Cys Gln Gly Gln Ala Cys Gln Asp Arg Trp Phe Ser
      980      985      990
Thr Pro Trp Ser Pro Cys Ser Arg Ser Cys Gln Gly Gly Thr Gln
      995      1000      1005
Thr Arg Glu Val Gln Cys Leu Ser Thr Asn Gln Thr Leu Ser Thr
      1010      1015      1020
Arg Cys Pro Pro Gln Leu Arg Pro Ser Arg Lys Arg Pro Cys Asn
      1025      1030      1035
Ser Gln Pro Cys Ser Gln Arg Pro Asp Asp Gln Cys Lys Asp Ser
      1040      1045      1050
Ser Pro His Cys Pro Leu Val Val Gln Ala Arg Leu Cys Val Tyr
      1055      1060      1065
Pro Tyr Tyr Thr Ala Thr Cys Cys Arg Ser Cys Ala His Val Leu
      1070      1075      1080
Glu Arg Ser Pro Gln Asp Pro Ser
      1085

```

```

<210> 20
<211> 436
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7504669CD1

```

```

<400> 20
Met Arg Ala Ala Arg Ala Ala Pro Leu Leu Gln Leu Leu Leu Leu
  1      5      10      15
Leu Gly Pro Trp Leu Glu Ala Ala Gly Val Ala Glu Ser Pro Leu
      20      25      30
Pro Ala Val Val Leu Ala Ile Leu Ala Arg Asn Ala Glu His Ser
      35      40      45
Leu Pro His Tyr Leu Gly Ala Leu Glu Arg Leu Asp Tyr Pro Arg
      50      55      60
Ala Arg Met Ala Leu Trp Cys Ala Thr Asp His Asn Val Asp Asn
      65      70      75
Thr Thr Glu Met Leu Gln Glu Trp Leu Ala Ala Val Gly Asp Asp
      80      85      90
Tyr Ala Ala Val Val Trp Arg Pro Glu Gly Glu Pro Arg Phe Tyr
      95      100      105
Pro Asp Glu Glu Gly Pro Lys His Trp Thr Lys Glu Arg His Gln
      110      115      120
Phe Leu Met Glu Leu Lys Gln Glu Ala Leu Thr Phe Ala Arg Asn
      125      130      135

```


| | | | |
|---|-----|-----|-----|
| Trp Gly Ala Asp Tyr Ile Leu Phe Ala Asp Thr Asp Asn Ile Leu | 140 | 145 | 150 |
| Thr Asn Asn Gln Thr Leu Arg Leu Leu Met Gly Gln Gly Leu Pro | 155 | 160 | 165 |
| Val Val Ala Pro Met Leu Asp Ser Gln Thr Tyr Tyr Ser Asn Phe | 170 | 175 | 180 |
| Trp Cys Gly Ile Thr Pro Gln Gly Tyr Tyr Arg Arg Thr Ala Glu | 185 | 190 | 195 |
| Tyr Phe Pro Thr Lys Asn Arg Gln Arg Arg Gly Cys Phe Arg Val | 200 | 205 | 210 |
| Pro Met Val His Ser Thr Phe Leu Ala Ser Leu Arg Ala Glu Gly | 215 | 220 | 225 |
| Ala Asp Gln Leu Ala Phe Tyr Pro Pro His Pro Asn Tyr Thr Trp | 230 | 235 | 240 |
| Pro Phe Asp Asp Ile Ile Val Phe Ala Tyr Ala Cys Gln Ala Ala | 245 | 250 | 255 |
| Gly Val Ser Val His Val Cys Asn Glu His Arg Tyr Gly Tyr Met | 260 | 265 | 270 |
| Asn Val Pro Val Lys Ser His Gln Gly Leu Glu Asp Glu Arg Val | 275 | 280 | 285 |
| Asn Phe Ile His Leu Ile Leu Glu Ala Leu Val Asp Gly Pro Arg | 290 | 295 | 300 |
| Met Gln Ala Ser Ala His Val Thr Arg Pro Ser Lys Arg Pro Ser | 305 | 310 | 315 |
| Lys Ile Gly Phe Asp Glu Val Phe Val Ile Ser Leu Ala Arg Arg | 320 | 325 | 330 |
| Pro Asp Arg Arg Glu Arg Met Leu Ala Ser Leu Trp Glu Met Glu | 335 | 340 | 345 |
| Ile Ser Gly Arg Val Val Asp Ala Val Asp Gly Trp Met Leu Asn | 350 | 355 | 360 |
| Ser Ser Ala Ile Arg Asn Leu Gly Val Asp Leu Leu Pro Gly Tyr | 365 | 370 | 375 |
| Gln Asp Pro Tyr Ser Gly Arg Thr Leu Thr Lys Gly Glu Val Gly | 380 | 385 | 390 |
| Cys Phe Leu Ser His Tyr Ser Ile Trp Glu Glu Arg Ala Val Gln | 395 | 400 | 405 |
| Gly Thr Leu Leu Ala Thr Gly Pro Gly Gly Leu Leu Arg Pro Ala | 410 | 415 | 420 |
| Pro Ala Arg Cys Pro Tyr Pro Leu Cys Arg Gly Arg Arg Val Ala | 425 | 430 | 435 |

Gln

<210> 21

<211> 70

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509266CD1

<400> 21

| | | | | |
|---|----|----|----|----|
| Met Gly Ile Glu Leu Leu Cys Leu Phe Phe Leu Phe Leu Gly Arg | 1 | 5 | 10 | 15 |
| Asn Asp His Val Gln Gly Gly Cys Ala Leu Gly Gly Ala Glu Thr | 20 | 25 | 30 | 35 |
| Cys Glu Asp Cys Leu Leu Ile Gly Pro Gln Cys Ala Trp Cys Ala | 40 | 45 | 50 | 55 |
| Gln Glu Asn Phe Thr His Pro Ser Gly Val Gly Glu Arg Trp Cys | 60 | 65 | 70 | |
| Ala Asp Ser Ala Gly Ala Cys Pro Pro Asp | | | | |

<210> 22
 <211> 715
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7509288CD1

<400> 22

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ile | Thr | Tyr | Lys | Val | Ala | Val | Pro | Trp | Glu | Val | Gln | Lys | Pro |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Val | Lys | Thr | Ala | Cys | Leu | Leu | Asp | Leu | Ser | Val | Pro | Gly | Val | Leu |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Arg | Arg | Ile | Leu | Leu | Ile | His | Leu | Glu | Leu | Ala | Lys | Gly | Gly | Ala |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Gln | Thr | Leu | Gln | Val | His | Val | Arg | Gln | Thr | Glu | Asp | Tyr | Pro | Val |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Asp | Leu | Tyr | Tyr | Leu | Met | Asp | Leu | Ser | Ala | Ser | Met | Asp | Asp | Asp |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Leu | Asn | Thr | Ile | Lys | Glu | Leu | Gly | Ser | Arg | Leu | Ser | Lys | Glu | Met |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Ser | Lys | Leu | Thr | Ser | Asn | Phe | Arg | Leu | Gly | Phe | Gly | Ser | Phe | Val |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Glu | Lys | Pro | Val | Ser | Pro | Phe | Val | Lys | Thr | Thr | Pro | Glu | Glu | Ile |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Ala | Asn | Pro | Cys | Ser | Ser | Ile | Pro | Tyr | Phe | Cys | Leu | Pro | Thr | Phe |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Gly | Phe | Lys | His | Ile | Leu | Pro | Leu | Thr | Asn | Asp | Ala | Glu | Arg | Phe |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Asn | Glu | Ile | Val | Lys | Asn | Gln | Lys | Ile | Ser | Ala | Asn | Ile | Asp | Thr |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Pro | Glu | Gly | Gly | Phe | Asp | Ala | Ile | Met | Gln | Ala | Ala | Val | Cys | Lys |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Glu | Lys | Ile | Gly | Trp | Arg | Asn | Asp | Ser | Leu | His | Leu | Leu | Val | Phe |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Val | Ser | Asp | Ala | Asp | Ser | His | Phe | Gly | Met | Asp | Ser | Lys | Leu | Ala |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Gly | Ile | Val | Ile | Pro | Asn | Asp | Gly | Leu | Cys | His | Leu | Asp | Ser | Lys |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Asn | Glu | Tyr | Ser | Met | Ser | Thr | Val | Leu | Glu | Tyr | Pro | Thr | Ile | Gly |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Gln | Leu | Ile | Asp | Lys | Leu | Val | Gln | Asn | Asn | Val | Leu | Leu | Ile | Phe |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Ala | Val | Thr | Gln | Glu | Gln | Val | His | Leu | Tyr | Glu | Asn | Tyr | Ala | Lys |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Leu | Ile | Pro | Gly | Ala | Thr | Val | Gly | Leu | Leu | Gln | Lys | Asp | Ser | Gly |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asn | Ile | Leu | Gln | Leu | Ile | Ile | Ser | Ala | Tyr | Glu | Glu | Leu | Arg | Ser |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Glu | Val | Glu | Leu | Glu | Val | Leu | Gly | Asp | Thr | Glu | Gly | Leu | Asn | Leu |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Ser | Phe | Thr | Ala | Ile | Cys | Asn | Asn | Gly | Thr | Leu | Phe | Gln | His | Gln |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Lys | Lys | Cys | Ser | His | Met | Lys | Val | Gly | Asp | Thr | Ala | Ser | Phe | Ser |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Val | Thr | Val | Asn | Ile | Pro | His | Cys | Glu | Arg | Ser | Arg | His | Ile | |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Ile | Ile | Lys | Pro | Val | Gly | Leu | Gly | Asp | Ala | Leu | Glu | Leu | Leu | Val |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Ser | Pro | Glu | Cys | Asn | Cys | Asp | Cys | Gln | Lys | Glu | Val | Glu | Val | Asn |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Ser | Ser | Lys | Cys | His | His | Gly | Asn | Gly | Ser | Phe | Gln | Cys | Gly | Val |

| | | | | | |
|-----------------|---------------------|---------------------|-----|--|-----|
| | 395 | | 400 | | 405 |
| Cys Ala Cys His | Pro Gly His Met Gly | Pro Arg Cys Glu Cys | Gly | | |
| | 410 | | 415 | | 420 |
| Glu Asp Met Leu | Ser Thr Asp Ser Cys | Lys Glu Ala Pro Asp | His | | |
| | 425 | | 430 | | 435 |
| Pro Ser Cys Ser | Gly Arg Gly Asp Cys | Tyr Cys Gly Gln Cys | Ile | | |
| | 440 | | 445 | | 450 |
| Cys His Leu Ser | Pro Tyr Gly Asn Ile | Tyr Gly Pro Tyr Cys | Gln | | |
| | 455 | | 460 | | 465 |
| Cys Asp Asn Phe | Ser Cys Val Arg His | Lys Gly Leu Leu Cys | Gly | | |
| | 470 | | 475 | | 480 |
| Gly Asn Gly Asp | Cys Asp Cys Gly Glu | Cys Val Cys Arg Ser | Gly | | |
| | 485 | | 490 | | 495 |
| Trp Thr Gly Glu | Tyr Cys Asn Cys Thr | Thr Ser Thr Asp Ser | Cys | | |
| | 500 | | 505 | | 510 |
| Val Ser Glu Asp | Gly Val Leu Cys Ser | Gly Arg Gly Asp Cys | Val | | |
| | 515 | | 520 | | 525 |
| Cys Gly Lys Cys | Val Cys Thr Asn Pro | Gly Ala Ser Gly Pro | Thr | | |
| | 530 | | 535 | | 540 |
| Cys Glu Arg Cys | Pro Thr Cys Gly Asp | Pro Cys Asn Ser Lys | Arg | | |
| | 545 | | 550 | | 555 |
| Ser Cys Ile Glu | Cys His Leu Ser Ala | Ala Gly Gln Ala Arg | Glu | | |
| | 560 | | 565 | | 570 |
| Glu Cys Val Asp | Lys Cys Lys Leu Ala | Gly Ala Thr Ile Ser | Glu | | |
| | 575 | | 580 | | 585 |
| Glu Glu Asp Phe | Ser Lys Asp Gly Ser | Val Ser Cys Ser Leu | Gln | | |
| | 590 | | 595 | | 600 |
| Gly Glu Asn Glu | Cys Leu Ile Thr Phe | Leu Ile Thr Thr Asp | Asn | | |
| | 605 | | 610 | | 615 |
| Glu Gly Lys Thr | Ile Ile His Ser Ile | Asn Glu Lys Asp Cys | Pro | | |
| | 620 | | 625 | | 630 |
| Lys Pro Pro Asn | Ile Pro Met Ile Met | Leu Gly Val Ser Leu | Ala | | |
| | 635 | | 640 | | 645 |
| Ile Leu Leu Ile | Gly Val Val Leu Leu | Cys Ile Trp Lys Leu | Leu | | |
| | 650 | | 655 | | 660 |
| Val Ser Phe His | Asp Arg Lys Glu Val | Ala Lys Phe Glu Ala | Glu | | |
| | 665 | | 670 | | 675 |
| Arg Ser Lys Ala | Lys Trp Gln Thr Gly | Thr Asn Pro Leu Tyr | Arg | | |
| | 680 | | 685 | | 690 |
| Gly Ser Thr Ser | Thr Phe Lys Asn Val | Thr Tyr Lys His Arg | Glu | | |
| | 695 | | 700 | | 705 |
| Lys Gln Lys Val | Asp Leu Ser Thr Asp | Cys | | | |
| | 710 | | 715 | | |

<210> 23

<211> 596

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510212CD1

<400> 23

| | | | | | |
|---------------------|---------------------|---------------------|--|--|----|
| Met Glu Arg Ala Ala | Pro Ser Arg Arg Val | Pro Leu Pro Leu Leu | | | |
| 1 | 5 | 10 | | | 15 |
| Leu Leu Gly Gly Leu | Ala Leu Leu Ala Ala | Gly Val Asp Ala Asp | | | |
| | 20 | 25 | | | 30 |
| Val Leu Leu Glu Ala | Cys Cys Ala Asp Gly | His Arg Met Ala Thr | | | |
| | 35 | 40 | | | 45 |
| His Gln Lys Asp Cys | Ser Leu Pro Tyr Ala | Thr Glu Ser Lys Glu | | | |
| | 50 | 55 | | | 60 |
| Cys Arg Met Val Gln | Glu Gln Cys Cys His | Ser Gln Leu Glu Glu | | | |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | 65 | | | | | 70 | | | | | 75 |
| Leu | His | Cys | Ala | Thr | Gly | Ile | Ser | Leu | Ala | Asn | Glu | Gln | Asp | Arg |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Cys | Ala | Thr | Pro | His | Gly | Asp | Asn | Ala | Ser | Leu | Glu | Ala | Thr | Phe |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Val | Lys | Arg | Cys | Cys | His | Cys | Cys | Leu | Leu | Gly | Arg | Ala | Ala | Gln |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Ala | Gln | Gly | Gln | Ser | Cys | Glu | Tyr | Ser | Leu | Met | Val | Gly | Tyr | Gln |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Cys | Gly | Gln | Val | Phe | Arg | Ala | Cys | Cys | Val | Lys | Ser | Gln | Glu | Thr |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Gly | Asp | Leu | Asp | Val | Gly | Gly | Leu | Gln | Glu | Thr | Asp | Lys | Ile | Ile |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Glu | Val | Glu | Glu | Glu | Gln | Glu | Asp | Pro | Tyr | Leu | Asn | Asp | Arg | Cys |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Arg | Gly | Gly | Gly | Pro | Cys | Lys | Gln | Gln | Cys | Arg | Asp | Thr | Gly | Asp |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Glu | Val | Val | Cys | Ser | Cys | Phe | Val | Gly | Tyr | Gln | Leu | Leu | Ser | Asp |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Gly | Val | Ser | Cys | Glu | Asp | Val | Asn | Glu | Cys | Ile | Thr | Gly | Ser | His |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Ser | Cys | Arg | Leu | Gly | Glu | Ser | Cys | Ile | Asn | Thr | Val | Gly | Ser | Phe |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Arg | Cys | Gln | Arg | Asp | Ser | Ser | Cys | Gly | Thr | Gly | Tyr | Glu | Leu | Thr |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Glu | Asp | Asn | Ser | Cys | Lys | Asp | Ile | Asp | Glu | Cys | Glu | Ser | Gly | Ile |
| | | | | 260 | | | | | 265 | | | | | 270 |
| His | Asn | Cys | Leu | Pro | Asp | Phe | Ile | Cys | Gln | Asn | Thr | Leu | Gly | Ser |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Phe | Arg | Cys | Arg | Pro | Lys | Leu | Gln | Cys | Lys | Ser | Gly | Phe | Ile | Gln |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Asp | Ala | Leu | Gly | Asn | Cys | Ile | Asp | Ile | Asn | Glu | Cys | Leu | Ser | Ile |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Ser | Ala | Pro | Cys | Pro | Ile | Gly | His | Thr | Cys | Ile | Asn | Thr | Glu | Gly |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Ser | Tyr | Thr | Cys | Gln | Lys | Asn | Val | Pro | Asn | Cys | Gly | Arg | Gly | Tyr |
| | | | | 335 | | | | | 340 | | | | | 345 |
| His | Leu | Asn | Glu | Glu | Gly | Thr | Arg | Cys | Val | Asp | Val | Asp | Glu | Cys |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Ala | Pro | Pro | Ala | Glu | Pro | Cys | Gly | Lys | Gly | His | Arg | Cys | Val | Asn |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Ser | Pro | Gly | Ser | Phe | Arg | Cys | Glu | Cys | Lys | Thr | Gly | Tyr | Tyr | Phe |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Asp | Gly | Ile | Ser | Arg | Met | Cys | Val | Asp | Val | Asn | Glu | Cys | Gln | Arg |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Tyr | Pro | Gly | Arg | Leu | Cys | Gly | His | Lys | Cys | Glu | Asn | Thr | Leu | Gly |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Ser | Tyr | Leu | Cys | Ser | Cys | Ser | Val | Gly | Phe | Arg | Leu | Ser | Val | Asp |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Gly | Arg | Ser | Cys | Glu | Glu | Ser | His | Lys | Cys | Glu | Asn | Thr | Leu | Gly |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Ser | Tyr | Leu | Cys | Ser | Cys | Ser | Val | Gly | Phe | Arg | Leu | Ser | Val | Asp |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Gly | Arg | Ser | Cys | Glu | Asp | Ile | Asn | Glu | Cys | Ser | Ser | Ser | Pro | Cys |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Ser | Gln | Glu | Cys | Ala | Asn | Val | Tyr | Gly | Ser | Tyr | Gln | Cys | Tyr | Cys |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Arg | Arg | Gly | Tyr | Gln | Leu | Ser | Asp | Val | Asp | Gly | Val | Thr | Cys | Glu |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Asp | Ile | Asp | Glu | Cys | Ala | Leu | Pro | Thr | Gly | Gly | His | Ile | Cys | Ser |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Tyr | Arg | Cys | Ile | Asn | Ile | Pro | Gly | Ser | Phe | Gln | Cys | Ser | Cys | Pro |
| | | | | 530 | | | | | 535 | | | | | 540 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ser | Ser | Gly | Tyr | Arg | Leu | Ala | Pro | Asn | Gly | Arg | Asn | Cys | Gln | Asp |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Ile | Asp | Glu | Cys | Val | Thr | Gly | Ile | His | Asn | Cys | Ser | Ile | Asn | Glu |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Thr | Cys | Phe | Asn | Ile | Gln | Gly | Gly | Phe | Arg | Cys | Leu | Ala | Phe | Glu |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Cys | Pro | Glu | Asn | Tyr | Arg | Arg | Ser | Ala | Ala | Thr | | | | |
| | | | | 590 | | | | | 595 | | | | | |

<210> 24

<211> 1762

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510504CD1

<400> 24

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Cys | Cys | Arg | Arg | Ala | Thr | Pro | Gly | Thr | Leu | Leu | Leu | Phe |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Leu | Ala | Phe | Leu | Leu | Leu | Ser | Ser | Arg | Thr | Ala | Arg | Ser | Glu | Glu |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Asp | Arg | Asp | Gly | Leu | Trp | Asp | Ala | Trp | Gly | Pro | Trp | Ser | Glu | Cys |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Ser | Arg | Thr | Cys | Gly | Gly | Gly | Ala | Ser | Tyr | Ser | Leu | Arg | Arg | Cys |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Leu | Ser | Ser | Lys | Ser | Cys | Glu | Gly | Arg | Asn | Ile | Arg | Tyr | Arg | Thr |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Cys | Ser | Asn | Val | Asp | Cys | Pro | Pro | Glu | Ala | Gly | Asp | Phe | Arg | Ala |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Gln | Gln | Cys | Ser | Ala | His | Asn | Asp | Val | Lys | His | His | Gly | Gln | Phe |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Tyr | Glu | Trp | Leu | Pro | Val | Ser | Asn | Asp | Pro | Asp | Asn | Pro | Cys | Ser |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Leu | Lys | Cys | Gln | Ala | Lys | Gly | Thr | Thr | Leu | Val | Val | Glu | Leu | Ala |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Pro | Lys | Val | Leu | Asp | Gly | Thr | Arg | Cys | Tyr | Thr | Glu | Ser | Leu | Asp |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Met | Cys | Ile | Ser | Gly | Leu | Cys | Gln | Ile | Val | Gly | Cys | Asp | His | Gln |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Leu | Gly | Ser | Thr | Val | Lys | Glu | Asp | Asn | Cys | Gly | Val | Cys | Asn | Gly |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Asp | Gly | Ser | Thr | Cys | Arg | Leu | Val | Arg | Gly | Gln | Tyr | Lys | Ser | Gln |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Leu | Ser | Ala | Thr | Lys | Ser | Asp | Asp | Thr | Val | Val | Ala | Ile | Pro | Tyr |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Gly | Ser | Arg | His | Ile | Arg | Leu | Val | Leu | Lys | Gly | Pro | Asp | His | Leu |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Tyr | Leu | Glu | Thr | Lys | Thr | Leu | Gln | Gly | Thr | Lys | Gly | Glu | Asn | Ser |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Leu | Ser | Ser | Thr | Gly | Thr | Phe | Leu | Val | Asp | Asn | Ser | Ser | Val | Asp |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Phe | Gln | Lys | Phe | Pro | Asp | Lys | Glu | Ile | Leu | Arg | Met | Ala | Gly | Pro |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Leu | Thr | Ala | Asp | Phe | Ile | Val | Lys | Ile | Arg | Asn | Ser | Gly | Ser | Ala |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asp | Ser | Thr | Val | Gln | Phe | Ile | Phe | Tyr | Gln | Pro | Ile | Ile | His | Arg |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Trp | Arg | Glu | Thr | Asp | Phe | Phe | Pro | Cys | Ser | Ala | Thr | Cys | Gly | Gly |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Gly | Tyr | Gln | Leu | Thr | Ser | Ala | Glu | Cys | Tyr | Asp | Leu | Arg | Ser | Asn |
| | | | | 320 | | | | | 325 | | | | | 330 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Arg | Val | Val | Ala | Asp | Gln | Tyr | Cys | His | Tyr | Tyr | Pro | Glu | Asn | Ile |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Lys | Pro | Lys | Pro | Lys | Leu | Gln | Glu | Cys | Asn | Leu | Asp | Pro | Cys | Pro |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Ala | Ser | Asp | Gly | Tyr | Lys | Gln | Ile | Met | Pro | Tyr | Asp | Leu | Tyr | His |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Pro | Leu | Pro | Arg | Trp | Glu | Ala | Thr | Pro | Trp | Thr | Ala | Cys | Ser | Ser |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Ser | Cys | Gly | Gly | Gly | Ile | Gln | Ser | Arg | Ala | Val | Ser | Cys | Val | Glu |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Glu | Asp | Ile | Gln | Gly | His | Val | Thr | Ser | Val | Glu | Glu | Trp | Lys | Cys |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Met | Tyr | Thr | Pro | Lys | Met | Pro | Ile | Ala | Gln | Pro | Cys | Asn | Ile | Phe |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Asp | Cys | Pro | Lys | Trp | Leu | Ala | Gln | Glu | Trp | Ser | Pro | Cys | Thr | Val |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Thr | Cys | Gly | Gln | Gly | Leu | Arg | Tyr | Arg | Val | Val | Leu | Cys | Ile | Asp |
| | | | | 455 | | | | | 460 | | | | | 465 |
| His | Arg | Gly | Met | His | Thr | Gly | Gly | Cys | Ser | Pro | Lys | Thr | Lys | Pro |
| | | | | 470 | | | | | 475 | | | | | 480 |
| His | Ile | Lys | Glu | Glu | Cys | Ile | Val | Pro | Thr | Pro | Cys | Tyr | Lys | Pro |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Lys | Glu | Lys | Leu | Pro | Val | Glu | Ala | Lys | Leu | Pro | Trp | Phe | Lys | Gln |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Ala | Gln | Glu | Leu | Glu | Glu | Gly | Ala | Ala | Val | Ser | Glu | Glu | Pro | Ser |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Phe | Ile | Pro | Glu | Ala | Trp | Ser | Ala | Cys | Thr | Val | Thr | Cys | Gly | Val |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Gly | Thr | Gln | Val | Arg | Ile | Val | Arg | Cys | Gln | Val | Leu | Leu | Ser | Phe |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Ser | Gln | Ser | Val | Ala | Asp | Leu | Pro | Ile | Asp | Glu | Cys | Glu | Gly | Pro |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Lys | Pro | Ala | Ser | Gln | Arg | Ala | Cys | His | Ala | Gly | Pro | Cys | Ser | Gly |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Glu | Ile | Pro | Glu | Phe | Asn | Pro | Asp | Glu | Thr | Asp | Gly | Leu | Phe | Gly |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Gly | Leu | Gln | Asp | Phe | Asp | Glu | Leu | Tyr | Asp | Trp | Glu | Tyr | Glu | Gly |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Phe | Thr | Lys | Cys | Ser | Glu | Ser | Cys | Gly | Gly | Gly | Val | Gln | Glu | Ala |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Val | Val | Ser | Cys | Leu | Asn | Lys | Gln | Thr | Arg | Glu | Pro | Ala | Glu | Glu |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Asn | Leu | Cys | Val | Thr | Ser | Arg | Arg | Pro | Pro | Gln | Leu | Leu | Lys | Ser |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Cys | Asn | Leu | Asp | Pro | Cys | Pro | Ala | Arg | Trp | Glu | Ile | Gly | Lys | Trp |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Ser | Pro | Cys | Ser | Leu | Thr | Cys | Gly | Val | Gly | Leu | Gln | Thr | Arg | Asp |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Val | Phe | Cys | Ser | His | Leu | Leu | Ser | Arg | Glu | Met | Asn | Glu | Thr | Val |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Ile | Leu | Ala | Asp | Glu | Leu | Cys | Arg | Gln | Pro | Lys | Pro | Ser | Thr | Val |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Gln | Ala | Cys | Asn | Arg | Phe | Asn | Cys | Pro | Pro | Ala | Trp | Tyr | Pro | Ala |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Gln | Trp | Gln | Pro | Cys | Ser | Arg | Thr | Cys | Gly | Gly | Gly | Val | Gln | Lys |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Arg | Glu | Val | Leu | Cys | Lys | Gln | Arg | Met | Ala | Asp | Gly | Ser | Phe | Leu |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Glu | Leu | Pro | Glu | Thr | Phe | Cys | Ser | Ala | Ser | Lys | Pro | Ala | Cys | Gln |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Gln | Ala | Cys | Lys | Lys | Asp | Asp | Cys | Pro | Ser | Glu | Trp | Leu | Leu | Ser |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Asp | Trp | Thr | Glu | Cys | Ser | Thr | Ser | Cys | Gly | Glu | Gly | Thr | Gln | Thr |

| | | | | | |
|-----------------|------|---------------------|------|-----------------|------|
| Arg Ser Ala Ile | 800 | Arg Lys Met Leu | 805 | Thr Gly Leu Ser | 810 |
| | 815 | | 820 | | 825 |
| Val Val Asn Ser | 830 | Leu Cys Pro Pro | 835 | Leu Pro Phe Ser | 840 |
| | 845 | | 850 | | 855 |
| Ile Arg Pro Cys | 860 | Met Leu Ala Thr Cys | 865 | Ala Arg Pro Gly | 870 |
| | 875 | | 880 | | 885 |
| Ser Thr Lys His | 890 | Arg Pro His Ile Ala | 895 | Ala Arg Lys Val | 900 |
| | 905 | | 910 | | 915 |
| Ile Gln Thr Arg | 920 | Arg Gln Arg Lys Leu | 925 | His Phe Val Val | 930 |
| | 935 | | 940 | | 945 |
| Phe Ala Tyr Leu | 950 | Leu Pro Lys Thr Ala | 955 | Val Val Leu Arg | 960 |
| | 965 | | 970 | | 975 |
| Ala Arg Arg Val | 980 | Arg Lys Pro Leu Ile | 985 | Thr Trp Glu Lys | 990 |
| | 995 | | 1000 | | 1005 |
| Gln His Leu Ile | 1010 | Ser Ser Thr His Val | 1015 | Thr Val Ala Pro | 1020 |
| | 1025 | | 1030 | | 1035 |
| Tyr Leu Lys Ile | 1040 | His Arg Leu Lys Pro | 1045 | Ser Asp Ala Gly | 1050 |
| | 1055 | | 1060 | | 1065 |
| Thr Cys Ser Ala | 1070 | Gly Pro Ala Arg Glu | 1075 | His Phe Val Ile | 1080 |
| | 1085 | | 1090 | | 1095 |
| Ile Gly Gly Asn | 1100 | Arg Lys Leu Val Ala | 1105 | Arg Pro Leu Ser | 1110 |
| | 1115 | | 1120 | | 1125 |
| Ser Glu Glu Glu | 1130 | Val Leu Ala Gly Arg | 1135 | Lys Gly Gly Pro | 1140 |
| | 1145 | | 1150 | | 1155 |
| Ala Leu Gln Thr | 1160 | His Lys His Gln Asn | 1165 | Gly Ile Phe Ser | 1170 |
| | 1175 | | 1180 | | 1185 |
| Ser Lys Ala Glu | 1190 | Lys Arg Gly Leu Ala | 1195 | Asn Pro Gly Ser | 1200 |
| | 1205 | | 1210 | | 1215 |
| Tyr Asp Asp Leu | 1220 | Val Ser Arg Leu Leu | 1225 | Gln Gly Gly Trp | 1230 |
| | 1235 | | 1240 | | 1245 |
| Gly Glu Leu Leu | 1250 | Ala Ser Trp Glu Ala | 1255 | Gln Asp Ser Ala | 1260 |
| | 1265 | | 1270 | | 1275 |
| Asn Thr Thr Ser | | Glu Asp Pro Gly Ala | | Glu Gln Val Leu | |
| | | | | | |
| His Leu Pro Phe | | Thr Met Val Thr | | Glu Gln Arg Arg | |
| | | | | | |
| Ile Leu Gly Asn | | Leu Ser Gln Gln | | Pro Glu Leu Arg | |
| | | | | | |
| Tyr Ser Lys His | | Leu Val Ala Gln | | Leu Ala Gln Glu | |
| | | | | | |
| Ser His Leu Glu | | His Gln Asp Thr | | Leu Leu Lys Pro | |
| | | | | | |
| Arg Thr Ser Pro | | Val Thr Leu Ser | | Pro His Lys His | |
| | | | | | |
| Phe Ser Ser Ser | | Leu Arg Thr Ser | | Ser Thr Gly Asp | |
| | | | | | |
| Gly Ser Arg Arg | | Pro His Arg Lys | | Pro Thr Ile Leu | |
| | | | | | |
| Ser Ala Ala Gln | | Gln Leu Ser Ala | | Ser Glu Val Val | |
| | | | | | |
| Gly Gln Thr Val | | Ala Leu Ala Ser | | Gly Thr Leu Ser | |
| | | | | | |
| His Cys Glu Ala | | Ile Gly His Pro | | Arg Pro Thr Ile | |
| | | | | | |
| Arg Asn Gly Glu | | Glu Val Gln Phe | | Ser Asp Arg Ile | |
| | | | | | |
| Pro Asp Asp Ser | | Leu Gln Ile Leu | | Ala Pro Val Glu | |
| | | | | | |
| Gly Phe Tyr Thr | | Cys Asn Ala Thr | | Asn Ala Leu Gly | |
| | | | | | |
| Val Ser Ile Ala | | Val Thr Leu Ala | | Gly Lys Pro Leu | |
| | | | | | |

| | | |
|---------------------|---------------------|---------------------|
| Ser Arg Met Thr Val | Ile Asn Thr Glu Lys | Pro Ala Val Thr Val |
| 1280 | 1285 | 1290 |
| Asp Ile Gly Ser Thr | Ile Lys Thr Val Gln | Gly Val Asn Val Thr |
| 1295 | 1300 | 1305 |
| Ile Asn Cys Gln Val | Ala Gly Val Pro Glu | Ala Glu Val Thr Trp |
| 1310 | 1315 | 1320 |
| Phe Arg Asn Lys Ser | Lys Leu Gly Ser Pro | His His Leu His Glu |
| 1325 | 1330 | 1335 |
| Gly Ser Leu Leu Leu | Thr Asn Val Ser Ser | Ser Asp Gln Gly Leu |
| 1340 | 1345 | 1350 |
| Tyr Ser Cys Arg Ala | Ala Asn Leu His Gly | Glu Leu Thr Glu Ser |
| 1355 | 1360 | 1365 |
| Thr Gln Leu Leu Ile | Leu Asp Pro Pro Gln | Val Pro Thr Gln Leu |
| 1370 | 1375 | 1380 |
| Glu Asp Ile Arg Ala | Leu Leu Ala Ala Thr | Gly Pro Asn Leu Pro |
| 1385 | 1390 | 1395 |
| Ser Val Leu Thr Ser | Pro Leu Gly Thr Gln | Leu Val Leu Asp Pro |
| 1400 | 1405 | 1410 |
| Gly Asn Ser Ala Leu | Leu Gly Cys Pro Ile | Lys Gly His Pro Val |
| 1415 | 1420 | 1425 |
| Pro Asn Ile Thr Trp | Phe His Gly Gly Gln | Pro Ile Val Thr Ala |
| 1430 | 1435 | 1440 |
| Thr Gly Leu Thr His | His Ile Leu Ala Ala | Gly Gln Ile Leu Gln |
| 1445 | 1450 | 1455 |
| Val Ala Asn Leu Ser | Gly Gly Ser Gln Gly | Glu Phe Ser Cys Leu |
| 1460 | 1465 | 1470 |
| Ala Gln Asn Glu Ala | Gly Val Leu Met Gln | Lys Ala Ser Leu Val |
| 1475 | 1480 | 1485 |
| Ile Gln Asp Tyr Trp | Trp Ser Val Asp Arg | Leu Ala Thr Cys Ser |
| 1490 | 1495 | 1500 |
| Ala Ser Cys Gly Asn | Arg Gly Val Gln Gln | Pro Arg Leu Arg Cys |
| 1505 | 1510 | 1515 |
| Leu Leu Asn Ser Thr | Glu Val Asn Pro Ala | His Cys Ala Gly Lys |
| 1520 | 1525 | 1530 |
| Val Arg Pro Ala Val | Gln Pro Ile Ala Cys | Asn Arg Arg Asp Cys |
| 1535 | 1540 | 1545 |
| Pro Ser Arg Trp Met | Val Thr Ser Trp Ser | Ala Cys Thr Arg Ser |
| 1550 | 1555 | 1560 |
| Cys Gly Gly Gly Val | Gln Thr Arg Arg Val | Thr Cys Gln Lys Leu |
| 1565 | 1570 | 1575 |
| Lys Ala Ser Gly Ile | Ser Thr Pro Val Ser | Asn Asp Met Cys Thr |
| 1580 | 1585 | 1590 |
| Gln Val Ala Lys Arg | Pro Val Asp Thr Gln | Ala Cys Asn Gln Gln |
| 1595 | 1600 | 1605 |
| Leu Cys Val Glu Trp | Ala Phe Ser Ser Trp | Gly Gln Cys Asn Gly |
| 1610 | 1615 | 1620 |
| Pro Cys Ile Gly Pro | His Leu Ala Val Gln | His Arg Gln Val Phe |
| 1625 | 1630 | 1635 |
| Cys Gln Thr Arg Asp | Gly Ile Thr Leu Pro | Ser Glu Gln Cys Ser |
| 1640 | 1645 | 1650 |
| Ala Leu Pro Arg Pro | Val Ser Thr Gln Asn | Cys Trp Ser Glu Ala |
| 1655 | 1660 | 1665 |
| Cys Ser Val His Trp | Arg Val Ser Leu Trp | Thr Leu Cys Thr Ala |
| 1670 | 1675 | 1680 |
| Thr Cys Gly Asn Tyr | Gly Phe Gln Ser Arg | Arg Val Glu Cys Val |
| 1685 | 1690 | 1695 |
| His Ala Arg Thr Asn | Lys Ala Val Pro Glu | His Leu Cys Ser Trp |
| 1700 | 1705 | 1710 |
| Gly Pro Arg Pro Ala | Asn Trp Gln Arg Cys | Asn Ile Thr Pro Cys |
| 1715 | 1720 | 1725 |
| Glu Asn Met Glu Cys | Arg Asp Thr Thr Arg | Tyr Cys Glu Lys Val |
| 1730 | 1735 | 1740 |
| Lys Gln Leu Lys Leu | Cys Gln Leu Ser Gln | Phe Lys Ser Arg Cys |

1745
 Cys Gly Thr Cys Gly Lys Ala
 1760
 1750
 1755
 <210> 25
 <211> 438
 <212> PRT
 <213> Homo sapiens
 <220>
 <221> misc_feature
 <223> Incyte ID No: 7510587CD1
 <400> 25
 Met Trp Thr Gly Tyr Lys Ile Leu Ile Phe Ser Tyr Leu Thr Thr
 1 5 10 15
 Glu Ile Trp Met Glu Lys Gln Tyr Leu Ser Gln Arg Glu Val Asp
 20 25 30
 Leu Glu Ala Tyr Phe Thr Arg Asn His Thr Val Leu Gln Gly Thr
 35 40 45
 Arg Phe Lys Arg Ala Ile Phe Gln Gly Gln Tyr Cys Arg Asn Phe
 50 55 60
 Gly Cys Cys Glu Asp Arg Asp Asp Gly Cys Val Thr Glu Phe Tyr
 65 70 75
 Ala Ala Asn Ala Leu Cys Tyr Cys Asp Lys Phe Cys Asp Arg Glu
 80 85 90
 Asn Ser Asp Cys Cys Pro Asp Tyr Lys Ser Phe Cys Arg Glu Glu
 95 100 105
 Lys Glu Trp Pro Pro His Thr Gln Pro Trp Tyr Pro Glu Gly Cys
 110 115 120
 Phe Lys Asp Gly Gln His Tyr Glu Glu Gly Ser Val Ile Lys Glu
 125 130 135
 Asn Cys Asn Ser Cys Thr Cys Ser Gly Gln Gln Trp Lys Cys Ser
 140 145 150
 Gln His Val Cys Leu Val Arg Pro Glu Leu Ile Glu Gln Val Asn
 155 160 165
 Lys Gly Asp Tyr Gly Trp Thr Ala Gln Asn Tyr Ser Gln Phe Trp
 170 175 180
 Gly Met Thr Leu Glu Asp Gly Phe Lys Phe Arg Leu Gly Thr Leu
 185 190 195
 Pro Pro Ser Pro Met Leu Leu Ser Met Asn Glu Met Thr Ala Ser
 200 205 210
 Leu Pro Ala Thr Thr Asp Leu Pro Glu Phe Phe Val Ala Ser Tyr
 215 220 225
 Lys Trp Pro Gly Trp Thr His Gly Pro Leu Asp Gln Lys Asn Cys
 230 235 240
 Ala Ala Ser Trp Ala Phe Ser Thr Ala Ser Val Ala Ala Asp Arg
 245 250 255
 Ile Ala Ile Gln Ser Lys Gly Arg Tyr Thr Ala Asn Leu Ser Pro
 260 265 270
 Gln Asn Leu Ile Ser Cys Cys Ala Lys Asn Arg His Gly Cys Asn
 275 280 285
 Ser Gly Ser Ile Asp Arg Ala Trp Trp Tyr Leu Arg Lys Arg Gly
 290 295 300
 Leu Val Ser His Ala Cys Tyr Pro Leu Phe Lys Asp Gln Asn Ala
 305 310 315
 Thr Asn Asn Gly Cys Ala Met Ala Ser Arg Ser Asp Gly Arg Gly
 320 325 330
 Lys Arg His Ala Thr Lys Pro Cys Pro Asn Asn Val Glu Lys Ser
 335 340 345
 Asn Arg Ile Tyr Gln Cys Ser Pro Pro Tyr Arg Val Ser Ser Asn
 350 355 360
 Glu Thr Glu Ile Met Lys Glu Ile Met Gln Asn Gly Pro Val Gln

| | | | | |
|-----------------|---------------------|-----|---------------------|-----|
| Ala Ile Met Gln | Val Arg Glu Asp Phe | 365 | 370 | 375 |
| | | | Phe His Tyr Lys Thr | Gly |
| Ile Tyr Arg His | Val Thr Ser Thr Asn | 380 | 385 | 390 |
| | | | Lys Glu Ser Glu Lys | Tyr |
| Arg Lys Leu Gln | Thr His Ala Val Lys | 395 | 400 | 405 |
| | | | Leu Thr Gly Leu Leu | Pro |
| Ile Pro Gly Glu | Ser His Gly Glu Arg | 410 | 415 | 420 |
| | | | Met Ala Ile Ser Gly | Phe |
| Phe Glu Glu | | 425 | 430 | 435 |

<210> 26
 <211> 401
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7510684CD1

<400> 26

| | | |
|---------------------|---------------------|---------------------|
| Met Trp Thr Gly Tyr | Lys Ile Leu Ile Phe | Ser Tyr Leu Thr Thr |
| 1 | 5 | 10 15 |
| Glu Ile Trp Met Glu | Lys Gln Tyr Leu Ser | Gln Arg Glu Val Asp |
| | 20 | 25 30 |
| Leu Glu Ala Tyr Phe | Thr Arg Asn His Thr | Val Leu Gln Gly Cys |
| | 35 | 40 45 |
| Phe Lys Asp Gly Gln | His Tyr Glu Glu Gly | Ser Val Ile Lys Glu |
| | 50 | 55 60 |
| Asn Cys Asn Ser Cys | Thr Cys Ser Gly Gln | Gln Trp Lys Cys Ser |
| | 65 | 70 75 |
| Gln His Val Cys Leu | Val Arg Ser Glu Leu | Ile Glu Gln Val Asn |
| | 80 | 85 90 |
| Lys Gly Asp Tyr Gly | Trp Thr Ala Gln Asn | Tyr Ser Gln Phe Trp |
| | 95 | 100 105 |
| Gly Met Thr Leu Glu | Asp Gly Phe Lys Phe | Arg Leu Gly Thr Leu |
| | 110 | 115 120 |
| Pro Pro Ser Pro Met | Leu Leu Ser Met Asn | Glu Met Thr Ala Ser |
| | 125 | 130 135 |
| Leu Pro Ala Thr Thr | Asp Leu Pro Glu Phe | Phe Val Ala Ser Tyr |
| | 140 | 145 150 |
| Lys Trp Pro Gly Trp | Thr His Gly Pro Leu | Asp Gln Lys Asn Cys |
| | 155 | 160 165 |
| Ala Ala Ser Trp Ala | Phe Ser Thr Ala Ser | Val Ala Ala Asp Arg |
| | 170 | 175 180 |
| Ile Ala Ile Gln Ser | Lys Gly Arg Tyr Thr | Ala Asn Leu Ser Pro |
| | 185 | 190 195 |
| Gln Asn Leu Ile Ser | Cys Cys Ala Lys Asn | Arg His Gly Cys Asn |
| | 200 | 205 210 |
| Ser Gly Ser Ile Asp | Arg Ala Trp Trp Tyr | Leu Arg Lys Arg Gly |
| | 215 | 220 225 |
| Leu Val Ser His Ala | Cys Tyr Pro Leu Phe | Lys Asp Gln Asn Ala |
| | 230 | 235 240 |
| Thr Asn Asn Gly Cys | Ala Met Ala Ser Arg | Ser Asp Gly Arg Gly |
| | 245 | 250 255 |
| Lys Arg His Ala Thr | Lys Pro Cys Pro Asn | Asn Val Glu Lys Ser |
| | 260 | 265 270 |
| Asn Arg Ile Tyr Gln | Cys Ser Pro Pro Tyr | Arg Val Ser Ser Asn |
| | 275 | 280 285 |
| Glu Thr Glu Ile Met | Lys Glu Ile Met Gln | Asn Gly Pro Val Gln |
| | 290 | 295 300 |
| Ala Ile Met Gln Val | Arg Glu Asp Phe Phe | His Tyr Lys Thr Gly |

| | | | | | |
|-----------------|-----|---------------------|-----|-------------------------|-----|
| Ile Tyr Arg His | 305 | Thr Ser Thr Asn | 310 | Glu Ser Glu Lys Tyr | 315 |
| | 320 | | 325 | | 330 |
| Arg Lys Leu Gln | 335 | Thr His Ala Val Lys | 340 | Leu Thr Gly Trp Gly Thr | 345 |
| | 350 | | 355 | | 360 |
| Leu Arg Gly Ala | 365 | Gln Gly Gln Lys Glu | 370 | Lys Phe Trp Ile Ala Ala | 375 |
| | 380 | | 385 | | 390 |
| Asn Ser Trp Gly | 395 | Lys Ser Trp Gly Glu | 400 | Asn Gly Tyr Phe Arg Ile | |
| | | | | | |
| Leu Arg Gly Val | | Asn Glu Ser Asp Ile | | Glu Lys Leu Ile Ile Ala | |
| | | | | | |
| Ala Trp Gly Gln | | Leu Thr Ser Ser Asp | | Glu Pro | |

<210> 27

<211> 1074

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510697CD1

<400> 27

| | | | |
|---------------------|-----------------|---------------------|-----|
| Met Glu Asn Trp Thr | Gly Arg Pro Trp | Leu Tyr Leu Leu Leu | Leu |
| 1 | 5 | 10 | 15 |
| Leu Ser Leu Pro Gln | Leu Cys Leu Asp | Gln Glu Val Leu Ser | Gly |
| | 20 | 25 | 30 |
| His Ser Leu Gln Thr | Pro Thr Glu Glu | Gly Gln Gly Pro Glu | Gly |
| | 35 | 40 | 45 |
| Val Trp Gly Pro Trp | Val Gln Trp Ala | Ser Cys Ser Gln Pro | Cys |
| | 50 | 55 | 60 |
| Gly Val Gly Val Gln | Arg Arg Ser Arg | Thr Cys Gln Leu Pro | Thr |
| | 65 | 70 | 75 |
| Val Gln Leu His Pro | Ser Leu Pro Leu | Pro Pro Arg Pro Pro | Arg |
| | 80 | 85 | 90 |
| His Pro Glu Ala Leu | Leu Pro Arg Gly | Gln Gly Pro Arg Pro | Gln |
| | 95 | 100 | 105 |
| Thr Ser Pro Glu Thr | Leu Pro Leu Tyr | Arg Thr Gln Ser Arg | Gly |
| | 110 | 115 | 120 |
| Arg Gly Gly Pro Leu | Arg Gly Pro Ala | Ser His Leu Gly Arg | Glu |
| | 125 | 130 | 135 |
| Glu Thr Gln Glu Ile | Arg Ala Ala Arg | Arg Ser Arg Leu Arg | Asp |
| | 140 | 145 | 150 |
| Pro Ile Lys Pro Gly | Met Phe Gly Tyr | Gly Arg Val Pro Phe | Ala |
| | 155 | 160 | 165 |
| Leu Pro Leu His Arg | Asn Arg Arg His | Pro Arg Ser Pro Pro | Arg |
| | 170 | 175 | 180 |
| Ser Glu Leu Ser Leu | Ile Ser Ser Arg | Gly Glu Glu Ala Ile | Pro |
| | 185 | 190 | 195 |
| Ser Pro Thr Pro Arg | Ala Glu Pro Phe | Ser Ala Asn Gly Ser | Pro |
| | 200 | 205 | 210 |
| Gln Thr Glu Leu Pro | Pro Thr Glu Leu | Ser Val His Thr Pro | Ser |
| | 215 | 220 | 225 |
| Pro Gln Ala Glu Pro | Leu Ser Pro Glu | Thr Ala Gln Thr Glu | Val |
| | 230 | 235 | 240 |
| Ala Pro Arg Thr Arg | Pro Ala Pro Leu | Arg His His Pro Arg | Ala |
| | 245 | 250 | 255 |
| Gln Ala Ser Gly Thr | Glu Pro Pro Ser | Pro Thr His Ser Leu | Gly |
| | 260 | 265 | 270 |
| Glu Gly Gly Phe Phe | Arg Ala Ser Pro | Gln Pro Arg Arg Pro | Ser |
| | 275 | 280 | 285 |
| Ser Gln Gly Trp Ala | Ser Pro Gln Val | Ala Gly Arg Arg Pro | Asp |

| | | | |
|---------------------|-----|---------------------|-----|
| Pro Phe Pro Ser | 290 | 295 | 300 |
| Val Pro Arg Gly Arg | 305 | Gly Gln Gln Gly Gln | Gly |
| Pro Trp Gly Thr | 320 | 310 | 315 |
| Asp Pro Gln His | 335 | 325 | 330 |
| Pro His Ala Ser | 350 | 340 | 345 |
| Ile Pro Arg Cys | 365 | 355 | 360 |
| Gln Ala Pro Cys | 380 | 370 | 375 |
| Cys Ala Ala Phe | 395 | 385 | 390 |
| Trp Glu Pro Phe | 410 | 400 | 405 |
| Asn Cys Arg Pro | 425 | 415 | 420 |
| Lys Val Gln Asp | 440 | 430 | 435 |
| Cys Val Ala Gly | 455 | 445 | 450 |
| Gly Ser Gly Arg | 470 | 460 | 465 |
| Asp Ser Thr Cys | 485 | 475 | 480 |
| Gly Pro Leu Gly | 500 | 490 | 495 |
| Leu Arg Leu Gln | 515 | 505 | 510 |
| Ala Leu Arg Gly | 530 | 520 | 525 |
| Ala Val Asp Pro | 545 | 535 | 540 |
| Arg Tyr Asn Arg | 560 | 550 | 555 |
| Ser Ala Glu Gly | 575 | 565 | 570 |
| Phe Gln Glu Glu | 590 | 580 | 585 |
| Ser Pro Pro Pro | 605 | 595 | 600 |
| Pro Gln Leu Gln | 620 | 610 | 615 |
| Pro Ala Pro Arg | 635 | 625 | 630 |
| Val Arg Ile Pro | 650 | 640 | 645 |
| Leu Gly Ser Pro | 665 | 655 | 660 |
| Cys Ser Ala Ser | 680 | 670 | 675 |
| Cys Ile Ser Arg | 695 | 685 | 690 |
| Ala Ala Gly Ala | 710 | 700 | 705 |
| Thr Pro Cys Pro | 725 | 715 | 720 |
| Ser Arg Ser Cys | 740 | 730 | 735 |
| Arg Gln Glu Phe | 755 | 745 | 750 |
| | | 760 | 765 |

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|------|
| Cys | Gly | His | Leu | Pro | Arg | Pro | Asn | Ile | Thr | Gln | Ser | Cys | Gln | Leu | 770 | 775 | 780 |
| Arg | Leu | Cys | Gly | His | Trp | Glu | Val | Gly | Ser | Pro | Trp | Ser | Gln | Cys | 785 | 790 | 795 |
| Ser | Val | Arg | Cys | Gly | Arg | Gly | Gln | Arg | Ser | Arg | Gln | Val | Arg | Cys | 800 | 805 | 810 |
| Val | Gly | Asn | Asn | Gly | Asp | Glu | Val | Ser | Glu | Gln | Glu | Cys | Ala | Ser | 815 | 820 | 825 |
| Gly | Pro | Pro | Gln | Pro | Pro | Ser | Arg | Glu | Ala | Cys | Asp | Met | Gly | Pro | 830 | 835 | 840 |
| Cys | Thr | Thr | Ala | Trp | Phe | His | Ser | Asp | Trp | Ser | Ser | Lys | Cys | Ser | 845 | 850 | 855 |
| Ala | Glu | Cys | Gly | Thr | Gly | Ile | Gln | Arg | Arg | Ser | Val | Val | Cys | Leu | 860 | 865 | 870 |
| Gly | Ser | Gly | Ala | Ala | Leu | Gly | Pro | Gly | Gln | Gly | Glu | Ala | Gly | Ala | 875 | 880 | 885 |
| Gly | Thr | Gly | Gln | Ser | Cys | Pro | Thr | Gly | Ser | Arg | Pro | Pro | Asp | Met | 890 | 895 | 900 |
| Arg | Ala | Cys | Ser | Leu | Gly | Pro | Cys | Glu | Arg | Thr | Trp | Arg | Trp | Tyr | 905 | 910 | 915 |
| Thr | Gly | Pro | Trp | Gly | Glu | Cys | Ser | Ser | Glu | Cys | Gly | Ser | Gly | Thr | 920 | 925 | 930 |
| Gln | Arg | Arg | Asp | Ile | Ile | Cys | Val | Ser | Lys | Leu | Gly | Thr | Glu | Phe | 935 | 940 | 945 |
| Asn | Val | Thr | Ser | Pro | Ser | Asn | Cys | Ser | His | Leu | Pro | Arg | Pro | Pro | 950 | 955 | 960 |
| Ala | Leu | Gln | Pro | Cys | Gln | Gly | Gln | Ala | Cys | Gln | Asp | Arg | Trp | Phe | 965 | 970 | 975 |
| Ser | Thr | Pro | Trp | Ser | Pro | Cys | Ser | Arg | Ser | Cys | Gln | Gly | Gly | Thr | 980 | 985 | 990 |
| Gln | Thr | Arg | Glu | Val | Gln | Cys | Leu | Ser | Thr | Asn | Gln | Thr | Leu | Ser | 995 | 1000 | 1005 |
| Thr | Arg | Cys | Pro | Pro | Gln | Leu | Arg | Pro | Ser | Arg | Lys | Arg | Pro | Cys | 1010 | 1015 | 1020 |
| Asn | Ser | Gln | Pro | Cys | Ser | Gln | Arg | Pro | Asp | Asp | Gln | Cys | Lys | Asp | 1025 | 1030 | 1035 |
| Ser | Ser | Pro | His | Cys | Pro | Leu | Val | Val | Gln | Ala | Arg | Leu | Cys | Val | 1040 | 1045 | 1050 |
| Tyr | Pro | Tyr | Tyr | Thr | Ala | Thr | Cys | Cys | Arg | Ser | Cys | Ala | His | Val | 1055 | 1060 | 1065 |
| Leu | Glu | Arg | Ser | Pro | Gln | Asp | Pro | Ser | | | | | | | 1070 | | |

<210> 28

<211> 1564

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7761337CD1

<400> 28

| | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|----|
| Met | Gly | Ala | Met | Thr | Gln | Leu | Leu | Ala | Gly | Val | Phe | Leu | Ala | Phe | 1 | 5 | 10 | 15 |
| Leu | Ala | Leu | Ala | Thr | Glu | Gly | Gly | Val | Leu | Lys | Lys | Val | Ile | Arg | 20 | 25 | 30 | 35 |
| His | Lys | Arg | Gln | Ser | Gly | Val | Asn | Ala | Thr | Leu | Pro | Glu | Glu | Asn | 40 | 45 | 50 | 55 |
| Gln | Pro | Val | Val | Phe | Asn | His | Val | Tyr | Asn | Ile | Lys | Leu | Pro | Val | 60 | 65 | 70 | 75 |
| Gly | Ser | Gln | Cys | Ser | Val | Asp | Leu | Glu | Ser | Ala | Ser | Gly | Glu | Lys | | | | |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Asp | Leu | Ala | Pro | Pro | Ser | Glu | Pro | Ser | Glu | Ser | Phe | Gln | Glu | His |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Thr | Val | Asp | Gly | Glu | Asn | Gln | Ile | Val | Phe | Thr | His | Arg | Ile | Asn |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Ile | Pro | Arg | Arg | Ala | Cys | Gly | Cys | Ala | Ala | Ala | Pro | Asp | Val | Lys |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Glu | Leu | Leu | Ser | Arg | Leu | Glu | Glu | Leu | Glu | Asn | Leu | Val | Ser | Ser |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Leu | Arg | Glu | Gln | Cys | Thr | Ala | Gly | Ala | Gly | Cys | Cys | Leu | Gln | Pro |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Ala | Thr | Gly | Arg | Leu | Asp | Thr | Arg | Pro | Phe | Cys | Ser | Gly | Arg | Gly |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Asn | Phe | Ser | Thr | Glu | Gly | Cys | Gly | Cys | Val | Cys | Glu | Pro | Gly | Trp |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Lys | Gly | Pro | Asn | Cys | Ser | Glu | Pro | Glu | Cys | Pro | Gly | Asn | Cys | His |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Leu | Arg | Gly | Arg | Cys | Ile | Asp | Gly | Gln | Cys | Ile | Cys | Asp | Asp | Gly |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Phe | Thr | Gly | Glu | Asp | Cys | Ser | Gln | Leu | Ala | Cys | Pro | Ser | Asp | Cys |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Asn | Asp | Gln | Gly | Lys | Cys | Val | Asn | Gly | Val | Cys | Ile | Cys | Phe | Glu |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Gly | Tyr | Ala | Gly | Ala | Asp | Cys | Ser | Arg | Glu | Ile | Cys | Pro | Val | Pro |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Cys | Ser | Glu | Glu | His | Gly | Thr | Cys | Val | Asp | Gly | Leu | Cys | Val | Cys |
| | | | | 260 | | | | | 265 | | | | | 270 |
| His | Asp | Gly | Phe | Ala | Gly | Asp | Asp | Cys | Asn | Lys | Pro | Leu | Cys | Leu |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asn | Asn | Cys | Tyr | Asn | Arg | Gly | Arg | Cys | Val | Glu | Asn | Glu | Cys | Val |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Cys | Asp | Glu | Gly | Phe | Thr | Gly | Glu | Asp | Cys | Ser | Glu | Leu | Ile | Cys |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Pro | Asn | Asp | Cys | Phe | Asp | Arg | Gly | Arg | Cys | Ile | Asn | Gly | Thr | Cys |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Tyr | Cys | Glu | Glu | Gly | Phe | Thr | Gly | Glu | Asp | Cys | Gly | Lys | Pro | Thr |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Cys | Pro | His | Ala | Cys | His | Thr | Gln | Gly | Arg | Cys | Glu | Glu | Gly | Gln |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Cys | Val | Cys | Asp | Glu | Gly | Phe | Ala | Gly | Leu | Asp | Cys | Ser | Glu | Lys |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Arg | Cys | Pro | Ala | Asp | Cys | His | Asn | Arg | Gly | Arg | Cys | Val | Asp | Gly |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Arg | Cys | Glu | Cys | Asp | Asp | Gly | Phe | Thr | Gly | Ala | Asp | Cys | Gly | Glu |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Leu | Lys | Cys | Pro | Asn | Gly | Cys | Ser | Gly | His | Gly | Arg | Cys | Val | Asn |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Gly | Gln | Cys | Val | Cys | Asp | Glu | Gly | Tyr | Thr | Gly | Glu | Asp | Cys | Ser |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Gln | Leu | Arg | Cys | Pro | Asn | Asp | Cys | His | Ser | Arg | Gly | Arg | Cys | Val |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Glu | Gly | Lys | Cys | Val | Cys | Glu | Gln | Gly | Phe | Lys | Gly | Tyr | Asp | Cys |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Ser | Asp | Met | Ser | Cys | Pro | Asn | Asp | Cys | His | Gln | His | Gly | Arg | Cys |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Val | Asn | Gly | Met | Cys | Val | Cys | Asp | Asp | Gly | Tyr | Thr | Gly | Glu | Asp |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Cys | Arg | Asp | Arg | Gln | Cys | Pro | Arg | Asp | Cys | Ser | Asn | Arg | Gly | Leu |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Cys | Val | Asp | Gly | Gln | Cys | Val | Cys | Glu | Asp | Gly | Phe | Thr | Gly | Pro |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Asp | Cys | Ala | Glu | Leu | Ser | Cys | Pro | Asn | Asp | Cys | His | Gly | Gln | Gly |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Arg | Cys | Val | Asn | Gly | Gln | Cys | Val | Cys | His | Glu | Gly | Phe | Met | Gly |

| | | | | | |
|-----------------|------|---------------------|------|---------------------|------|
| Lys Asp Cys Lys | 545 | Glu Gln Arg Cys Pro | 550 | Ser Asp Cys His Gly | 555 |
| Gly Arg Cys Val | 560 | Asp Gly Gln Cys Ile | 565 | Cys His Glu Gly Phe | 570 |
| Gly Leu Asp Cys | 575 | Gly Gln His Ser Cys | 580 | Pro Ser Asp Cys Asn | 585 |
| Leu Gly Gln Cys | 590 | Val Ser Gly Arg Cys | 595 | Ile Cys Asn Glu Gly | 600 |
| Ser Gly Glu Asp | 605 | Cys Ser Glu Val Ser | 610 | Pro Pro Lys Asp Leu | 615 |
| Val Thr Glu Val | 620 | Thr Glu Glu Thr Val | 625 | Asn Leu Ala Trp Asp | 630 |
| Glu Met Arg Val | 635 | Thr Glu Tyr Leu Val | 640 | Val Tyr Thr Pro Thr | 645 |
| Glu Gly Gly Leu | 650 | Glu Met Gln Phe Arg | 655 | Val Pro Gly Asp Gln | 660 |
| Ser Thr Ile Ile | 665 | Gln Glu Leu Glu Pro | 670 | Gly Val Glu Tyr Phe | 675 |
| Arg Val Phe Ala | 680 | Ile Leu Glu Asn Lys | 685 | Lys Ser Ile Pro Val | 690 |
| Ala Arg Val Ala | 695 | Thr Tyr Leu Pro Ala | 700 | Pro Glu Gly Leu Lys | 705 |
| Lys Ser Ile Lys | 710 | Glu Thr Ser Val Glu | 715 | Val Glu Trp Asp Pro | 720 |
| Asp Ile Ala Phe | 725 | Glu Thr Trp Glu Ile | 730 | Ile Phe Arg Asn Met | 735 |
| Lys Glu Asp Glu | 740 | Gly Glu Ile Thr Lys | 745 | Ser Leu Arg Arg Pro | 750 |
| Thr Ser Tyr Arg | 755 | Gln Thr Gly Leu Ala | 760 | Pro Gly Gln Glu Tyr | 765 |
| Ile Ser Leu His | 770 | Ile Val Lys Asn Asn | 775 | Thr Arg Gly Pro Gly | 780 |
| Lys Arg Val Thr | 785 | Thr Thr Arg Leu Asp | 790 | Ala Pro Ser Gln Ile | 795 |
| Val Lys Asp Val | 800 | Thr Asp Thr Thr Ala | 805 | Leu Ile Thr Trp Phe | 810 |
| Pro Leu Ala Glu | 815 | Ile Asp Gly Ile Glu | 820 | Leu Thr Tyr Gly Ile | 825 |
| Asp Val Pro Gly | 830 | Asp Arg Thr Thr Ile | 835 | Leu Thr Glu Asp Glu | 840 |
| Asn Gln Tyr Ser | 845 | Ile Gly Asn Leu Lys | 850 | Pro Asp Thr Glu Tyr | 855 |
| Val Ser Leu Ile | 860 | Ser Arg Arg Gly Asp | 865 | Met Ser Ser Asn Pro | 870 |
| Lys Glu Thr Phe | 875 | Thr Thr Gly Leu Asp | 880 | Ala Pro Arg Asn Leu | 885 |
| Arg Val Ser Gln | 890 | Thr Asp Asn Ser Ile | 895 | Thr Leu Glu Trp Arg | 900 |
| Gly Lys Ala Ala | 905 | Ile Asp Ser Tyr Arg | 910 | Ile Lys Tyr Ala Pro | 915 |
| Ser Gly Gly Asp | 920 | His Ala Glu Val Asp | 925 | Val Pro Lys Ser Gln | 930 |
| Ala Thr Thr Lys | 935 | Thr Thr Leu Thr Gly | 940 | Leu Arg Pro Gly Thr | 945 |
| Tyr Gly Ile Gly | 950 | Val Ser Ala Val Lys | 955 | Glu Asp Lys Glu Ser | 960 |
| Pro Ala Thr Ile | 965 | Asn Ala Ala Thr Glu | 970 | Leu Asp Thr Pro Lys | 975 |
| Leu Gln Val Ser | 980 | Glu Thr Ala Glu Thr | 985 | Ser Leu Thr Leu Leu | 990 |
| Lys Thr Pro Leu | 995 | Ala Lys Phe Asp Arg | 1000 | Tyr Arg Leu Asn Tyr | 1005 |
| | 1010 | | 1015 | | 1020 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|
| Leu | Pro | Thr | Gly | Gln | Trp | Val | Gly | Val | Gln | Leu | Pro | Arg | Asn | Thr |
| | | | | 1025 | | | | | 1030 | | | | | 1035 |
| Thr | Ser | Tyr | Val | Leu | Arg | Gly | Leu | Glu | Pro | Gly | Gln | Glu | Tyr | Asn |
| | | | | 1040 | | | | | 1045 | | | | | 1050 |
| Val | Leu | Leu | Thr | Ala | Glu | Lys | Gly | Arg | His | Lys | Ser | Lys | Pro | Ala |
| | | | | 1055 | | | | | 1060 | | | | | 1065 |
| Arg | Val | Lys | Ala | Ser | Thr | Ala | Met | Gly | Ser | Pro | Lys | Glu | Val | Ile |
| | | | | 1070 | | | | | 1075 | | | | | 1080 |
| Phe | Ser | Asp | Ile | Thr | Glu | Asn | Ser | Ala | Thr | Val | Ser | Trp | Arg | Ala |
| | | | | 1085 | | | | | 1090 | | | | | 1095 |
| Pro | Thr | Ala | Gln | Val | Glu | Ser | Phe | Arg | Ile | Thr | Tyr | Val | Pro | Ile |
| | | | | 1100 | | | | | 1105 | | | | | 1110 |
| Thr | Gly | Gly | Thr | Pro | Ser | Met | Val | Thr | Val | Asp | Gly | Thr | Lys | Thr |
| | | | | 1115 | | | | | 1120 | | | | | 1125 |
| Gln | Thr | Arg | Leu | Val | Lys | Leu | Ile | Pro | Gly | Val | Glu | Tyr | Leu | Val |
| | | | | 1130 | | | | | 1135 | | | | | 1140 |
| Ser | Ile | Ile | Ala | Met | Lys | Gly | Phe | Glu | Glu | Ser | Glu | Pro | Val | Ser |
| | | | | 1145 | | | | | 1150 | | | | | 1155 |
| Gly | Ser | Phe | Thr | Thr | Ala | Leu | Asp | Gly | Pro | Ser | Gly | Leu | Val | Thr |
| | | | | 1160 | | | | | 1165 | | | | | 1170 |
| Ala | Asn | Ile | Thr | Asp | Ser | Glu | Ala | Leu | Ala | Arg | Trp | Gln | Pro | Ala |
| | | | | 1175 | | | | | 1180 | | | | | 1185 |
| Ile | Ala | Thr | Val | Asp | Ser | Tyr | Val | Ile | Ser | Tyr | Thr | Gly | Glu | Lys |
| | | | | 1190 | | | | | 1195 | | | | | 1200 |
| Val | Pro | Glu | Ile | Thr | Arg | Thr | Val | Ser | Gly | Asn | Thr | Val | Glu | Tyr |
| | | | | 1205 | | | | | 1210 | | | | | 1215 |
| Ala | Leu | Thr | Asp | Leu | Glu | Pro | Ala | Thr | Glu | Tyr | Thr | Leu | Arg | Ile |
| | | | | 1220 | | | | | 1225 | | | | | 1230 |
| Phe | Ala | Glu | Lys | Gly | Pro | Gln | Lys | Ser | Ser | Thr | Ile | Thr | Ala | Lys |
| | | | | 1235 | | | | | 1240 | | | | | 1245 |
| Phe | Thr | Thr | Asp | Leu | Asp | Ser | Pro | Arg | Asp | Leu | Thr | Ala | Thr | Glu |
| | | | | 1250 | | | | | 1255 | | | | | 1260 |
| Val | Gln | Ser | Glu | Thr | Ala | Leu | Leu | Thr | Trp | Arg | Pro | Pro | Arg | Ala |
| | | | | 1265 | | | | | 1270 | | | | | 1275 |
| Ser | Val | Thr | Gly | Tyr | Leu | Leu | Val | Tyr | Glu | Ser | Val | Asp | Gly | Thr |
| | | | | 1280 | | | | | 1285 | | | | | 1290 |
| Val | Lys | Glu | Val | Ile | Val | Gly | Pro | Asp | Thr | Thr | Ser | Tyr | Ser | Leu |
| | | | | 1295 | | | | | 1300 | | | | | 1305 |
| Ala | Asp | Leu | Ser | Pro | Ser | Thr | His | Tyr | Thr | Ala | Lys | Ile | Gln | Ala |
| | | | | 1310 | | | | | 1315 | | | | | 1320 |
| Leu | Asn | Gly | Pro | Leu | Arg | Ser | Asn | Met | Ile | Gln | Thr | Ile | Phe | Thr |
| | | | | 1325 | | | | | 1330 | | | | | 1335 |
| Thr | Ile | Gly | Leu | Leu | Tyr | Pro | Phe | Pro | Lys | Asp | Cys | Ser | Gln | Ala |
| | | | | 1340 | | | | | 1345 | | | | | 1350 |
| Met | Leu | Asn | Gly | Asp | Thr | Thr | Ser | Gly | Leu | Tyr | Thr | Ile | Tyr | Leu |
| | | | | 1355 | | | | | 1360 | | | | | 1365 |
| Asn | Gly | Asp | Lys | Ala | Gln | Ala | Leu | Glu | Val | Phe | Cys | Asp | Met | Thr |
| | | | | 1370 | | | | | 1375 | | | | | 1380 |
| Ser | Asp | Gly | Gly | Gly | Trp | Ile | Val | Phe | Leu | Arg | Arg | Lys | Asn | Gly |
| | | | | 1385 | | | | | 1390 | | | | | 1395 |
| Arg | Glu | Asn | Phe | Tyr | Gln | Asn | Trp | Lys | Ala | Tyr | Ala | Ala | Gly | Phe |
| | | | | 1400 | | | | | 1405 | | | | | 1410 |
| Gly | Asp | Arg | Arg | Glu | Glu | Phe | Trp | Leu | Gly | Leu | Asp | Asn | Leu | Asn |
| | | | | 1415 | | | | | 1420 | | | | | 1425 |
| Lys | Ile | Thr | Ala | Gln | Gly | Gln | Tyr | Glu | Leu | Arg | Val | Asp | Leu | Arg |
| | | | | 1430 | | | | | 1435 | | | | | 1440 |
| Asp | His | Gly | Glu | Thr | Ala | Phe | Ala | Val | Tyr | Asp | Lys | Phe | Ser | Val |
| | | | | 1445 | | | | | 1450 | | | | | 1455 |
| Gly | Asp | Ala | Lys | Thr | Arg | Tyr | Lys | Leu | Lys | Val | Glu | Gly | Tyr | Ser |
| | | | | 1460 | | | | | 1465 | | | | | 1470 |
| Gly | Thr | Ala | Gly | Asp | Ser | Met | Ala | Tyr | His | Asn | Gly | Arg | Ser | Phe |
| | | | | 1475 | | | | | 1480 | | | | | 1485 |
| Ser | Thr | Phe | Asp | Lys | Asp | Thr | Asp | Ser | Ala | Ile | Thr | Asn | Cys | Ala |

| | | |
|---------------------|---------------------|---------------------|
| 1490 | 1495 | 1500 |
| Leu Ser Tyr Lys Gly | Ala Phe Trp Tyr Arg | Asn Cys His Arg Val |
| 1505 | 1510 | 1515 |
| Asn Leu Met Gly Arg | Tyr Gly Asp Asn Asn | His Ser Gln Gly Val |
| 1520 | 1525 | 1530 |
| Asn Trp Phe His Trp | Lys Gly His Glu His | Ser Ile Gln Phe Ala |
| 1535 | 1540 | 1545 |
| Glu Met Lys Leu Arg | Pro Ser Asn Phe Arg | Asn Leu Glu Gly Arg |
| 1550 | 1555 | 1560 |
| Arg Lys Arg Ala | | |

<210> 29
 <211> 834
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503666CD1

<400> 29

| | | |
|---------------------|---------------------|-------------------------|
| Met Gly Ser Trp Ala | Leu Leu Trp Pro | Pro Leu Leu Phe Thr Gly |
| 1 | 5 | 10 15 |
| Leu Leu Val Arg Pro | Pro Gly Thr Met Ala | Gln Ala Gln Tyr Cys |
| 20 | 25 | 30 |
| Ser Val Asn Lys Asp | Ile Phe Glu Val Glu | Asn Thr Asn Val |
| 35 | 40 | 45 |
| Thr Glu Pro Leu Val | Asp Ile His Val Pro | Glu Gly Gln Glu Val |
| 50 | 55 | 60 |
| Thr Leu Gly Ala Leu | Ser Thr Pro Phe Ala | Phe Arg Ile Gln Gly |
| 65 | 70 | 75 |
| Asn Gln Leu Phe Leu | Asn Val Thr Pro Asp | Tyr Glu Glu Lys Ser |
| 80 | 85 | 90 |
| Leu Leu Glu Ala Gln | Leu Leu Cys Gln Ser | Gly Gly Thr Leu Val |
| 95 | 100 | 105 |
| Thr Gln Leu Arg Val | Phe Val Ser Val Leu | Asp Val Asn Asp Asn |
| 110 | 115 | 120 |
| Ala Pro Glu Phe Pro | Phe Lys Thr Lys Glu | Ile Arg Val Glu Glu |
| 125 | 130 | 135 |
| Asp Thr Lys Val Asn | Ser Thr Val Ile Pro | Glu Thr Gln Leu Gln |
| 140 | 145 | 150 |
| Ala Glu Asp Arg Asp | Lys Asp Asp Ile Leu | Phe Tyr Thr Leu Gln |
| 155 | 160 | 165 |
| Glu Met Thr Ala Gly | Ala Ser Asp Tyr Phe | Ser Leu Val Ser Val |
| 170 | 175 | 180 |
| Asn Arg Pro Ala Leu | Arg Leu Asp Arg Pro | Leu Asp Phe Tyr Glu |
| 185 | 190 | 195 |
| Arg Pro Asn Met Thr | Phe Trp Leu Leu Val | Arg Asp Thr Pro Gly |
| 200 | 205 | 210 |
| Glu Asn Val Glu Pro | Ser His Thr Ala Thr | Ala Thr Leu Val Leu |
| 215 | 220 | 225 |
| Asn Val Val Pro Ala | Asp Leu Arg Pro Pro | Trp Phe Leu Pro Cys |
| 230 | 235 | 240 |
| Thr Phe Ser Asp Gly | Tyr Val Cys Ile Gln | Ala Gln Tyr His Gly |
| 245 | 250 | 255 |
| Ala Val Pro Thr Gly | His Ile Leu Pro Ser | Pro Leu Val Leu Arg |
| 260 | 265 | 270 |
| Pro Gly Pro Ile Tyr | Ala Glu Asp Gly Asp | Arg Gly Ile Asn Gln |
| 275 | 280 | 285 |
| Pro Ile Ile Tyr Ser | Ile Phe Arg Asp Ser | Gly Asn Leu Thr Val |
| 290 | 295 | 300 |
| Ala Arg Ser Val Pro | Ser Pro Met Thr Phe | Leu Leu Leu Val Lys |

| | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | 305 | | | | | 310 | | | | 315 |
| Gly | Gln | Gln | Ala | Asp | Leu | Ala | Arg | Tyr | Ser | Val | Thr | Gln | Val |
| | | | | 320 | | | | | 325 | | | | 330 |
| Val | Glu | Ala | Val | Ala | Ala | Ala | Gly | Ser | Pro | Pro | Arg | Phe | Pro |
| | | | | 335 | | | | | 340 | | | | 345 |
| Arg | Leu | Tyr | Arg | Gly | Thr | Val | Ala | Arg | Gly | Ala | Gly | Ala | Gly |
| | | | | 350 | | | | | 355 | | | | 360 |
| Val | Val | Lys | Asp | Ala | Ala | Ala | Pro | Ser | Gln | Pro | Leu | Arg | Ile |
| | | | | 365 | | | | | 370 | | | | 375 |
| Ala | Gln | Asp | Pro | Glu | Phe | Ser | Asp | Leu | Asn | Ser | Ala | Ile | Thr |
| | | | | 380 | | | | | 385 | | | | 390 |
| Arg | Ile | Thr | Asn | His | Ser | His | Phe | Arg | Met | Glu | Gly | Glu | Val |
| | | | | 395 | | | | | 400 | | | | 405 |
| Leu | Thr | Thr | Thr | Thr | Leu | Ala | Gln | Ala | Gly | Ala | Phe | Tyr | Ala |
| | | | | 410 | | | | | 415 | | | | 420 |
| Val | Glu | Ala | His | Asn | Thr | Val | Thr | Ser | Gly | Thr | Ala | Thr | Thr |
| | | | | 425 | | | | | 430 | | | | 435 |
| Ile | Glu | Ile | Gln | Val | Ser | Glu | Gln | Glu | Pro | Pro | Ser | Thr | Asp |
| | | | | 440 | | | | | 445 | | | | 450 |
| Pro | Pro | Ser | Pro | Glu | Ala | Gly | Gly | Thr | Thr | Gly | Pro | Trp | Thr |
| | | | | 455 | | | | | 460 | | | | 465 |
| Thr | Thr | Ser | Glu | Val | Pro | Arg | Pro | Pro | Glu | Pro | Ser | Gln | Gly |
| | | | | 470 | | | | | 475 | | | | 480 |
| Ser | Thr | Thr | Ser | Ser | Gly | Gly | Gly | Thr | Gly | Pro | His | Pro | Pro |
| | | | | 485 | | | | | 490 | | | | 495 |
| Gly | Thr | Thr | Leu | Arg | Pro | Pro | Thr | Ser | Ser | Thr | Pro | Gly | Gly |
| | | | | 500 | | | | | 505 | | | | 510 |
| Pro | Gly | Ala | Glu | Asn | Ser | Thr | Ser | His | Gln | Pro | Ala | Thr | Pro |
| | | | | 515 | | | | | 520 | | | | 525 |
| Gly | Asp | Thr | Ala | Gln | Thr | Pro | Lys | Pro | Gly | Thr | Ser | Gln | Pro |
| | | | | 530 | | | | | 535 | | | | 540 |
| Pro | Pro | Gly | Val | Gly | Thr | Ser | Thr | Ser | His | Gln | Pro | Ala | Thr |
| | | | | 545 | | | | | 550 | | | | 555 |
| Ser | Gly | Gly | Thr | Ala | Gln | Thr | Pro | Glu | Pro | Gly | Thr | Ser | Gln |
| | | | | 560 | | | | | 565 | | | | 570 |
| Met | Pro | Pro | Ser | Met | Gly | Thr | Ser | Thr | Ser | His | Gln | Pro | Ala |
| | | | | 575 | | | | | 580 | | | | 585 |
| Pro | Gly | Gly | Gly | Thr | Ala | Gln | Thr | Pro | Glu | Ala | Gly | Thr | Ser |
| | | | | 590 | | | | | 595 | | | | 600 |
| Pro | Met | Pro | Pro | Gly | Met | Gly | Thr | Ser | Thr | Ser | His | Gln | Pro |
| | | | | 605 | | | | | 610 | | | | 615 |
| Thr | Pro | Gly | Gly | Gly | Thr | Ala | Gln | Thr | Pro | Glu | Pro | Gly | Thr |
| | | | | 620 | | | | | 625 | | | | 630 |
| Gln | Pro | Met | Pro | Leu | Ser | Lys | Ser | Thr | Pro | Ser | Ser | Gly | Gly |
| | | | | 635 | | | | | 640 | | | | 645 |
| Pro | Ser | Glu | Asp | Lys | Arg | Phe | Ser | Val | Val | Asp | Met | Ala | Ala |
| | | | | 650 | | | | | 655 | | | | 660 |
| Gly | Gly | Val | Leu | Gly | Ala | Leu | Leu | Leu | Leu | Ala | Leu | Leu | Gly |
| | | | | 665 | | | | | 670 | | | | 675 |
| Ala | Val | Leu | Val | His | Lys | His | Tyr | Gly | Pro | Arg | Leu | Lys | Cys |
| | | | | 680 | | | | | 685 | | | | 690 |
| Ser | Gly | Lys | Ala | Pro | Glu | Pro | Gln | Pro | Gln | Gly | Phe | Asp | Asn |
| | | | | 695 | | | | | 700 | | | | 705 |
| Ala | Phe | Leu | Pro | Asp | His | Lys | Ala | Asn | Trp | Ala | Pro | Val | Pro |
| | | | | 710 | | | | | 715 | | | | 720 |
| Pro | Thr | His | Asp | Pro | Lys | Pro | Ala | Glu | Ala | Pro | Met | Pro | Ala |
| | | | | 725 | | | | | 730 | | | | 735 |
| Pro | Ala | Pro | Pro | Gly | Pro | Ala | Ser | Pro | Gly | Gly | Ala | Pro | Glu |
| | | | | 740 | | | | | 745 | | | | 750 |
| Pro | Ala | Ala | Ala | Arg | Ala | Gly | Gly | Ser | Pro | Thr | Ala | Val | Arg |
| | | | | 755 | | | | | 760 | | | | 765 |
| Ile | Leu | Thr | Lys | Glu | Arg | Arg | Pro | Glu | Gly | Gly | Tyr | Lys | Ala |
| | | | | 770 | | | | | 775 | | | | 780 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Trp | Phe | Gly | Glu | Asp | Ile | Gly | Thr | Glu | Ala | Asp | Val | Val | Val | Leu |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Asn | Ala | Pro | Thr | Leu | Asp | Val | Asp | Gly | Ala | Ser | Asp | Ser | Gly | Ser |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Gly | Asp | Glu | Gly | Glu | Gly | Ala | Gly | Arg | Gly | Gly | Gly | Pro | Tyr | Asp |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Ala | Pro | Gly | Gly | Asp | Asp | Ser | Tyr | Ile | | | | | | |
| | | | | 830 | | | | | | | | | | |

<210> 30

<211> 814

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503668CD1

<400> 30

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Gly | Ser | Trp | Ala | Leu | Leu | Trp | Pro | Pro | Leu | Leu | Phe | Thr | Gly |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Leu | Leu | Val | Arg | Pro | Pro | Gly | Thr | Met | Ala | Gln | Ala | Gln | Tyr | Cys |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Ser | Val | Asn | Lys | Asp | Ile | Phe | Glu | Val | Glu | Glu | Asn | Thr | Asn | Val |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Thr | Glu | Pro | Leu | Val | Asp | Ile | His | Val | Pro | Glu | Gly | Gln | Glu | Val |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Thr | Leu | Gly | Ala | Leu | Ser | Thr | Pro | Phe | Ala | Phe | Arg | Ile | Gln | Gly |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Asn | Gln | Leu | Phe | Leu | Asn | Val | Thr | Pro | Asp | Tyr | Glu | Glu | Lys | Ser |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Leu | Leu | Glu | Ala | Gln | Leu | Leu | Cys | Gln | Ser | Gly | Gly | Thr | Leu | Asp |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Thr | Lys | Val | Asn | Ser | Thr | Val | Ile | Pro | Glu | Thr | Gln | Leu | Gln | Ala |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Glu | Asp | Arg | Asp | Lys | Asp | Asp | Ile | Leu | Phe | Tyr | Thr | Leu | Gln | Glu |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Met | Thr | Ala | Gly | Ala | Ser | Asp | Tyr | Phe | Ser | Leu | Val | Ser | Val | Asn |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Arg | Pro | Ala | Leu | Arg | Leu | Asp | Arg | Pro | Leu | Asp | Phe | Tyr | Glu | Arg |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Pro | Asn | Met | Thr | Phe | Trp | Leu | Leu | Val | Arg | Asp | Thr | Pro | Gly | Glu |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Asn | Val | Glu | Pro | Ser | His | Thr | Ala | Thr | Ala | Thr | Leu | Val | Leu | Asn |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Val | Val | Pro | Ala | Asp | Leu | Arg | Pro | Pro | Trp | Phe | Leu | Pro | Cys | Thr |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Phe | Ser | Asp | Gly | Tyr | Val | Cys | Ile | Gln | Ala | Gln | Tyr | His | Gly | Ala |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Val | Pro | Thr | Gly | His | Ile | Leu | Pro | Ser | Pro | Leu | Val | Leu | Arg | Pro |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Gly | Pro | Ile | Tyr | Ala | Glu | Asp | Gly | Asp | Arg | Gly | Ile | Asn | Gln | Pro |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Ile | Ile | Tyr | Ser | Ile | Phe | Arg | Gly | Asn | Val | Asn | Gly | Thr | Phe | Ile |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Ile | His | Pro | Asp | Ser | Gly | Asn | Leu | Thr | Val | Ala | Arg | Ser | Val | Pro |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Ser | Pro | Met | Thr | Phe | Leu | Leu | Leu | Val | Lys | Gly | Gln | Gln | Ala | Asp |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Leu | Ala | Arg | Tyr | Ser | Val | Thr | Gln | Val | Thr | Val | Glu | Ala | Val | Ala |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Ala | Ala | Gly | Ser | Pro | Pro | Arg | Phe | Pro | Gln | Arg | Leu | Tyr | Arg | Gly |
| | | | | 320 | | | | | 325 | | | | | 330 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Thr | Val | Ala | Arg | Gly | Ala | Gly | Ala | Gly | Val | Val | Val | Lys | Asp | Ala |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Ala | Ala | Pro | Ser | Gln | Pro | Leu | Arg | Ile | Gln | Ala | Gln | Asp | Pro | Glu |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Phe | Ser | Asp | Leu | Asn | Ser | Ala | Ile | Thr | Tyr | Arg | Ile | Thr | Asn | His |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Ser | His | Phe | Arg | Met | Glu | Gly | Glu | Val | Val | Leu | Thr | Thr | Thr | Thr |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Leu | Ala | Gln | Ala | Gly | Ala | Phe | Tyr | Ala | Glu | Val | Glu | Ala | His | Asn |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Thr | Val | Thr | Ser | Gly | Thr | Ala | Thr | Thr | Val | Ile | Glu | Ile | Gln | Val |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Ser | Glu | Gln | Glu | Pro | Pro | Ser | Thr | Asp | Val | Pro | Pro | Ser | Pro | Glu |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Ala | Gly | Gly | Thr | Thr | Gly | Pro | Trp | Thr | Ser | Thr | Thr | Ser | Glu | Val |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Pro | Arg | Pro | Pro | Glu | Pro | Ser | Gln | Gly | Pro | Ser | Thr | Thr | Ser | Ser |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Gly | Gly | Gly | Thr | Gly | Pro | His | Pro | Pro | Ser | Gly | Thr | Thr | Leu | Arg |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Pro | Pro | Thr | Ser | Ser | Thr | Pro | Gly | Gly | Pro | Pro | Gly | Ala | Glu | Asn |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Ser | Thr | Ser | His | Gln | Pro | Ala | Thr | Pro | Gly | Gly | Asp | Thr | Ala | Gln |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Thr | Pro | Lys | Pro | Gly | Thr | Ser | Gln | Pro | Met | Pro | Pro | Gly | Val | Gly |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Thr | Ser | Thr | Ser | His | Gln | Pro | Ala | Thr | Pro | Ser | Gly | Gly | Thr | Ala |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Gln | Thr | Pro | Glu | Pro | Gly | Thr | Ser | Gln | Pro | Met | Pro | Pro | Ser | Met |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Gly | Thr | Ser | Thr | Ser | His | Gln | Pro | Ala | Thr | Pro | Gly | Gly | Gly | Thr |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Ala | Gln | Thr | Pro | Glu | Ala | Gly | Thr | Ser | Gln | Pro | Met | Pro | Pro | Gly |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Met | Gly | Thr | Ser | Thr | Ser | His | Gln | Pro | Thr | Thr | Pro | Gly | Gly | Gly |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Thr | Ala | Gln | Thr | Pro | Glu | Pro | Gly | Thr | Ser | Gln | Pro | Met | Pro | Leu |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Ser | Lys | Ser | Thr | Pro | Ser | Ser | Gly | Gly | Gly | Pro | Ser | Glu | Asp | Lys |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Arg | Phe | Ser | Val | Val | Asp | Met | Ala | Ala | Leu | Gly | Gly | Val | Leu | Gly |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Ala | Leu | Leu | Leu | Leu | Ala | Leu | Leu | Gly | Leu | Ala | Val | Leu | Val | His |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Lys | His | Tyr | Gly | Pro | Arg | Leu | Lys | Cys | Cys | Ser | Gly | Lys | Ala | Pro |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Glu | Pro | Gln | Pro | Gln | Gly | Phe | Asp | Asn | Gln | Ala | Phe | Leu | Pro | Asp |
| | | | | 680 | | | | | 685 | | | | | 690 |
| His | Lys | Ala | Asn | Trp | Ala | Pro | Val | Pro | Ser | Pro | Thr | His | Asp | Pro |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Lys | Pro | Ala | Glu | Ala | Pro | Met | Pro | Ala | Glu | Pro | Ala | Pro | Pro | Gly |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Pro | Ala | Ser | Pro | Gly | Gly | Ala | Pro | Glu | Pro | Pro | Ala | Ala | Ala | Arg |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Ala | Gly | Gly | Ser | Pro | Thr | Ala | Val | Arg | Ser | Ile | Leu | Thr | Lys | Glu |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Arg | Arg | Pro | Glu | Gly | Gly | Tyr | Lys | Ala | Val | Trp | Phe | Gly | Glu | Asp |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Ile | Gly | Thr | Glu | Ala | Asp | Val | Val | Val | Leu | Asn | Ala | Pro | Thr | Leu |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Asp | Val | Asp | Gly | Ala | Ser | Asp | Ser | Gly | Ser | Gly | Asp | Glu | Gly | Glu |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Gly | Ala | Gly | Arg | Gly | Gly | Gly | Pro | Tyr | Asp | Ala | Pro | Gly | Gly | Asp |

800 805 810
 Asp Ser Tyr Ile

<210> 31
 <211> 807
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503672CD1

<400> 31
 Met Gly Ser Trp Ala Leu Leu Trp Pro Pro Leu Leu Phe Thr Gly
 1 5 10 15
 Leu Leu Val Arg Pro Pro Gly Thr Met Ala Gln Ala Gln Tyr Cys
 20 25 30
 Ser Val Asn Lys Asp Ile Phe Glu Val Glu Glu Asn Thr Asn Val
 35 40 45
 Thr Glu Pro Leu Val Asp Ile His Val Pro Glu Gly Gln Glu Val
 50 55 60
 Thr Leu Gly Ala Leu Ser Thr Pro Phe Ala Phe Arg Ile Gln Gly
 65 70 75
 Asn Gln Leu Phe Leu Asn Val Thr Pro Asp Tyr Glu Glu Lys Ser
 80 85 90
 Leu Leu Glu Ala Gln Leu Leu Cys Gln Ser Gly Gly Thr Leu Val
 95 100 105
 Thr Gln Leu Arg Val Phe Val Ser Val Leu Asp Val Asn Asp Asn
 110 115 120
 Ala Pro Glu Phe Pro Phe Lys Thr Lys Glu Ile Arg Val Glu Glu
 125 130 135
 Asp Thr Lys Val Asn Ser Thr Val Ile Pro Glu Thr Gln Leu Gln
 140 145 150
 Ala Glu Asp Arg Asp Lys Asp Asp Ile Leu Phe Tyr Thr Leu Gln
 155 160 165
 Glu Met Thr Ala Gly Ala Ser Asp Tyr Phe Ser Leu Val Ser Val
 170 175 180
 Asn Arg Pro Ala Leu Arg Leu Asp Arg Pro Leu Asp Phe Tyr Glu
 185 190 195
 Arg Pro Asn Met Thr Phe Trp Leu Leu Val Arg Asp Thr Pro Gly
 200 205 210
 Glu Asn Val Glu Pro Ser His Thr Ala Thr Ala Thr Leu Val Leu
 215 220 225
 Asn Val Val Pro Ala Asp Leu Arg Pro Pro Trp Phe Leu Pro Cys
 230 235 240
 Thr Phe Ser Asp Gly Tyr Val Cys Ile Gln Ala Gln Tyr His Gly
 245 250 255
 Ala Val Pro Thr Gly His Ile Leu Pro Ser Pro Leu Val Leu Arg
 260 265 270
 Pro Gly Pro Ile Tyr Ala Glu Asp Gly Asp Arg Gly Ile Asn Gln
 275 280 285
 Pro Ile Ile Tyr Ser Ile Phe Arg Gly Asn Val Asn Gly Thr Phe
 290 295 300
 Ile Ile His Pro Asp Ser Gly Asn Leu Thr Val Ala Arg Ser Val
 305 310 315
 Pro Ser Pro Met Thr Phe Leu Leu Leu Val Lys Gly Gln Gln Ala
 320 325 330
 Asp Leu Ala Arg Tyr Ser Val Thr Gln Val Thr Val Glu Ala Val
 335 340 345
 Ala Ala Ala Gly Ser Pro Pro Arg Phe Pro Gln Arg Leu Tyr Arg
 350 355 360
 Gly Thr Val Ala Arg Gly Ala Gly Ala Gly Val Val Val Lys Asp

| | | | | | |
|-----------------|-----|---------------------|-----|-------------------------|-----|
| Ala Ala Ala Pro | 365 | Ser Gln Pro Leu Arg | 370 | Gln Ala Gln Asp | 375 |
| Glu Phe Ser Asp | 380 | Leu Asn Ser Ala Ile | 385 | Tyr Arg Ile Thr | 390 |
| His Ser His Phe | 395 | Arg Met Glu Gly Glu | 400 | Val Val Leu Thr Thr | 405 |
| Thr Leu Ala Gln | 410 | Ala Gly Ala Phe Tyr | 415 | Glu Val Glu Ala His | 420 |
| Asn Thr Val Thr | 425 | Ser Gly Thr Ala Thr | 430 | Thr Val Ile Glu Ile Gln | 435 |
| Val Ser Glu Gln | 440 | Glu Pro Pro Ser Thr | 445 | Asp Val Pro Pro Ser Pro | 450 |
| Glu Ala Gly Gly | 455 | Thr Thr Gly Pro Trp | 460 | Ser Thr Thr Ser Glu | 465 |
| Val Pro Arg Pro | 470 | Pro Glu Pro Ser Gln | 475 | Thr Thr Thr Ser Thr | 480 |
| Ser Gly Gly Gly | 485 | Thr Gly Pro His Pro | 490 | Gly Pro Ser Thr Thr Ser | 495 |
| Arg Pro Pro Thr | 500 | Ser Ser Thr Pro Gly | 505 | Pro Ser Gly Thr Thr Leu | 510 |
| Asn Ser Thr Ser | 515 | His Gln Pro Ala Thr | 520 | Gly Pro Pro Gly Ala Glu | 525 |
| Gln Thr Pro Lys | 530 | Pro Gly Thr Ser Gln | 535 | Pro Gly Gly Asp Thr Ala | 540 |
| Gly Thr Ser Thr | 545 | Ser His Gln Pro Ala | 550 | Thr Pro Ser Gly Gly Thr | 555 |
| Ala Gln Thr Pro | 560 | Glu Pro Gly Thr Ser | 565 | Pro Met Pro Pro Gly Ser | 570 |
| Met Gly Thr Ser | 575 | Thr Ser His Gln Pro | 580 | Ala Thr Pro Gly Gly Gly | 585 |
| Thr Ala Gln Thr | 590 | Pro Glu Ala Gly Thr | 595 | Ser Gln Pro Met Pro Pro | 600 |
| Gly Gly Gly Pro | 605 | Ser Glu Asp Lys Arg | 610 | Phe Ser Val Val Asp Met | 615 |
| Ala Ala Leu Gly | 620 | Gly Val Leu Gly Ala | 625 | Leu Leu Leu Leu Ala Leu | 630 |
| Leu Gly Leu Ala | 635 | Val Leu Val His Lys | 640 | His Tyr Gly Pro Arg Leu | 645 |
| Lys Cys Cys Ser | 650 | Gly Lys Ala Pro Glu | 655 | Pro Gln Pro Gln Gly Phe | 660 |
| Asp Asn Gln Ala | 665 | Phe Leu Pro Asp His | 670 | Lys Ala Asn Trp Ala Pro | 675 |
| Val Pro Ser Pro | 680 | Thr His Asp Pro Lys | 685 | Pro Ala Glu Ala Pro Met | 690 |
| Pro Ala Glu Pro | 695 | Ala Pro Pro Gly Pro | 700 | Ala Ser Pro Gly Gly Ala | 705 |
| Pro Glu Pro Pro | 710 | Ala Ala Ala Arg Ala | 715 | Gly Gly Ser Pro Thr Ala | 720 |
| Val Arg Ser Ile | 725 | Leu Thr Lys Glu Arg | 730 | Arg Pro Glu Gly Gly Tyr | 735 |
| Lys Ala Val Trp | 740 | Phe Gly Glu Asp Ile | 745 | Gly Thr Glu Ala Asp Val | 750 |
| Val Val Leu Asn | 755 | Ala Pro Thr Leu Asp | 760 | Val Asp Gly Ala Ser Asp | 765 |
| Ser Gly Ser Gly | 770 | Asp Glu Gly Glu Gly | 775 | Ala Gly Arg Gly Gly Gly | 780 |
| Pro Tyr Asp Ala | 785 | Pro Gly Gly Asp Asp | 790 | Ser Tyr Ile | 795 |
| | 800 | | 805 | | |

<210> 32
 <211> 1232
 <212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6039650CD1

<400> 32

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Met | Cys | Glu | Val | Met | Pro | Thr | Ile | Ser | Glu | Ala | Glu | Gly | Pro |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Pro | Gly | Gly | Gly | Gly | Gly | His | Gly | Ser | Gly | Ser | Pro | Ser | Gln | Pro |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Asp | Ala | Asp | Ser | His | Phe | Glu | Gln | Leu | Met | Val | Ser | Met | Leu | Glu |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Glu | Arg | Asp | Arg | Leu | Leu | Asp | Thr | Leu | Arg | Glu | Thr | Gln | Glu | Thr |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Leu | Ala | Leu | Thr | Gln | Gly | Lys | Leu | His | Glu | Val | Gly | His | Glu | Arg |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Asp | Ser | Leu | Gln | Arg | Gln | Leu | Asn | Thr | Ala | Leu | Pro | Gln | Glu | Phe |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Ala | Ala | Leu | Thr | Lys | Glu | Leu | Asn | Val | Cys | Arg | Glu | Gln | Leu | Leu |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Glu | Arg | Glu | Glu | Glu | Ile | Ala | Glu | Leu | Lys | Ala | Glu | Arg | Asn | Asn |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Thr | Arg | Leu | Leu | Leu | Glu | His | Leu | Glu | Cys | Leu | Val | Ser | Arg | His |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Glu | Arg | Ser | Leu | Arg | Met | Thr | Val | Val | Lys | Arg | Gln | Ala | Gln | Ser |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Pro | Ala | Gly | Val | Ser | Ser | Glu | Val | Glu | Val | Leu | Lys | Ala | Leu | Lys |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Ser | Leu | Phe | Glu | His | His | Lys | Ala | Leu | Asp | Glu | Lys | Val | Arg | Glu |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Arg | Leu | Arg | Val | Ala | Leu | Glu | Arg | Cys | Ser | Leu | Leu | Glu | Glu | Glu |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Leu | Gly | Ala | Thr | His | Lys | Glu | Leu | Met | Ile | Leu | Lys | Glu | Gln | Asn |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Asn | Gln | Lys | Lys | Thr | Leu | Thr | Asp | Gly | Val | Leu | Asp | Ile | Asn | His |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Glu | Gln | Glu | Asn | Thr | Pro | Ser | Thr | Ser | Gly | Lys | Arg | Ser | Ser | Asp |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Gly | Ser | Leu | Ser | His | Glu | Glu | Asp | Leu | Ala | Lys | Val | Ile | Glu | Leu |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Gln | Glu | Ile | Ile | Ser | Lys | Gln | Ser | Arg | Glu | Gln | Ser | Gln | Met | Lys |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Glu | Arg | Leu | Ala | Ser | Leu | Ser | Ser | His | Val | Thr | Glu | Leu | Glu | Glu |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asp | Leu | Asp | Thr | Ala | Arg | Lys | Asp | Leu | Ile | Lys | Ser | Glu | Glu | Met |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Asn | Thr | Lys | Leu | Gln | Arg | Asp | Val | Arg | Glu | Ala | Met | Ala | Gln | Lys |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Glu | Asp | Met | Glu | Glu | Arg | Ile | Thr | Thr | Leu | Glu | Lys | Arg | Tyr | Leu |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Ala | Ala | Gln | Arg | Glu | Ala | Thr | Ser | Val | His | Asp | Leu | Asn | Asp | Lys |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Leu | Glu | Asn | Glu | Ile | Ala | Asn | Lys | Asp | Ser | Met | His | Arg | Gln | Thr |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Glu | Asp | Lys | Asn | Arg | Gln | Leu | Gln | Glu | Arg | Leu | Glu | Leu | Ala | Glu |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Gln | Lys | Leu | Gln | Gln | Thr | Leu | Arg | Lys | Ala | Glu | Thr | Leu | Pro | Glu |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Val | Glu | Ala | Glu | Leu | Ala | Gln | Arg | Val | Ala | Ala | Leu | Ser | Lys | Ala |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Glu | Glu | Arg | His | Gly | Asn | Ile | Glu | Glu | Arg | Leu | Arg | Gln | Met | Glu |
| | | | | 410 | | | | | 415 | | | | | 420 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Gln | Leu | Glu | Glu | Lys | Asn | Arg | Glu | Leu | Gln | Arg | Ala | Arg | Gln |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Arg | Glu | Lys | Met | Asn | Glu | Glu | His | Asn | Lys | Arg | Leu | Ser | Asp | Thr |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Val | Asp | Lys | Leu | Leu | Ser | Glu | Ser | Asn | Glu | Arg | Leu | Gln | Leu | His |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Leu | Lys | Glu | Arg | Met | Ala | Ala | Leu | Glu | Asp | Lys | Asn | Ser | Leu | Leu |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Arg | Glu | Val | Glu | Ser | Ala | Lys | Lys | Gln | Leu | Glu | Glu | Thr | Gln | His |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Asp | Lys | Asp | Gln | Leu | Val | Leu | Asn | Ile | Glu | Ala | Leu | Arg | Ala | Glu |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Leu | Asp | His | Met | Arg | Leu | Arg | Gly | Ala | Ser | Leu | His | His | Gly | Arg |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Pro | His | Leu | Gly | Ser | Val | Pro | Asp | Phe | Arg | Phe | Pro | Met | Ala | Asp |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Gly | His | Thr | Asp | Ser | Tyr | Ser | Thr | Ser | Ala | Val | Leu | Arg | Arg | Pro |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Gln | Lys | Gly | Arg | Leu | Ala | Ala | Leu | Arg | Asp | Glu | Pro | Ser | Lys | Val |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Gln | Thr | Leu | Asn | Glu | Gln | Asp | Trp | Glu | Arg | Ala | Gln | Gln | Ala | Ser |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Val | Leu | Ala | Asn | Val | Ala | Gln | Ala | Phe | Glu | Ser | Asp | Ala | Asp | Val |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Ser | Asp | Gly | Glu | Asp | Asp | Arg | Asp | Thr | Leu | Leu | Ser | Ser | Val | Asp |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Leu | Leu | Ser | Pro | Ser | Gly | Gln | Ala | Asp | Ala | His | Thr | Leu | Ala | Met |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Met | Leu | Gln | Glu | Gln | Leu | Asp | Ala | Ile | Asn | Lys | Glu | Ile | Arg | Leu |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Ile | Gln | Glu | Glu | Lys | Glu | Asn | Thr | Glu | Gln | Arg | Ala | Glu | Glu | Ile |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Glu | Ser | Arg | Val | Gly | Ser | Gly | Ser | Leu | Asp | Asn | Leu | Gly | Arg | Phe |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Arg | Ser | Met | Ser | Ser | Ile | Pro | Pro | Tyr | Pro | Ala | Ser | Ser | Leu | Ala |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Ser | Ser | Ser | Pro | Pro | Gly | Ser | Gly | Arg | Ser | Thr | Pro | Arg | Arg | Ile |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Pro | His | Ser | Pro | Ala | Arg | Glu | Val | Asp | Arg | Leu | Gly | Val | Met | Thr |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Leu | Leu | Pro | Pro | Ser | Arg | Glu | Glu | Val | Arg | Asp | Asp | Lys | Thr | Thr |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Ile | Lys | Cys | Glu | Thr | Ser | Pro | Pro | Ser | Ser | Pro | Arg | Ala | Leu | Arg |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Leu | Asp | Arg | Leu | His | Lys | Gly | Ala | Leu | His | Thr | Val | Ser | His | Glu |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Asp | Ile | Arg | Asp | Ile | Arg | Asn | Ser | Thr | Gly | Ser | Gln | Asp | Gly | Pro |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Val | Ser | Asn | Pro | Ser | Ser | Ser | Asn | Ser | Ser | Gln | Asp | Ser | Leu | His |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Lys | Ala | Pro | Lys | Lys | Lys | Gly | Ile | Lys | Ser | Ser | Ile | Gly | Arg | Leu |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Phe | Gly | Lys | Lys | Glu | Lys | Gly | Arg | Pro | Gly | Gln | Thr | Gly | Lys | Glu |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Ala | Leu | Gly | Gln | Ala | Gly | Val | Ser | Glu | Thr | Asp | Asn | Ser | Ser | Gln |
| | | | | 830 | | | | | 835 | | | | | 840 |
| Asp | Ala | Leu | Gly | Leu | Ser | Lys | Leu | Gly | Gly | Gln | Ala | Glu | Lys | Asn |
| | | | | 845 | | | | | 850 | | | | | 855 |
| Arg | Lys | Leu | Gln | Lys | Lys | His | Glu | Leu | Leu | Glu | Glu | Ala | Arg | Arg |
| | | | | 860 | | | | | 865 | | | | | 870 |
| Gln | Gly | Leu | Pro | Phe | Ala | Gln | Trp | Asp | Gly | Pro | Thr | Val | Val | Val |
| | | | | 875 | | | | | 880 | | | | | 885 |
| Trp | Leu | Glu | Leu | Trp | Val | Gly | Met | Pro | Ala | Trp | Tyr | Val | Ala | Ala |

| | | | | | |
|---|------|--|------|--|------|
| | 890 | | 895 | | 900 |
| Cys Arg Ala Asn Val Lys Ser Gly Ala Ile Met Ser Ala Leu Ser | | | | | |
| | 905 | | 910 | | 915 |
| Asp Thr Glu Ile Gln Arg Glu Ile Gly Ile Ser Asn Pro Leu His | | | | | |
| | 920 | | 925 | | 930 |
| Arg Leu Lys Leu Arg Leu Ala Ile Gln Glu Ile Met Ser Leu Thr | | | | | |
| | 935 | | 940 | | 945 |
| Ser Pro Ser Ala Pro Thr Ser Arg Thr Thr Gly Asn Val | | | | | |
| | 950 | | 955 | | 960 |
| Trp Leu Thr His Glu Glu Met Glu Thr Leu Ala Ala Thr Pro Gln | | | | | |
| | 965 | | 970 | | 975 |
| Thr Glu Asp Glu Glu Gly Ser Trp Ala Gln Thr Leu Ala Tyr Gly | | | | | |
| | 980 | | 985 | | 990 |
| Asp Met Asn His Glu Trp Ile Gly Asn Glu Trp Leu Pro Ser Leu | | | | | |
| | 995 | | 1000 | | 1005 |
| Gly Leu Pro Gln Tyr Arg Ser Tyr Phe Met Glu Cys Leu Val Asp | | | | | |
| | 1010 | | 1015 | | 1020 |
| Ala Arg Met Leu Asp His Leu Thr Lys Lys Asp Leu Arg Gly Gln | | | | | |
| | 1025 | | 1030 | | 1035 |
| Leu Lys Met Val Asp Ser Phe His Arg Asn Ser Phe Gln Cys Gly | | | | | |
| | 1040 | | 1045 | | 1050 |
| Ile Met Cys Leu Arg Arg Leu Asn Tyr Asp Arg Lys Glu Leu Glu | | | | | |
| | 1055 | | 1060 | | 1065 |
| Arg Lys Arg Glu Glu Ser Gln Ser Glu Ile Lys Asp Val Leu Val | | | | | |
| | 1070 | | 1075 | | 1080 |
| Trp Ser Asn Asp Arg Val Ile Arg Trp Ile Leu Ser Ile Gly Leu | | | | | |
| | 1085 | | 1090 | | 1095 |
| Lys Glu Tyr Ala Asn Asn Leu Ile Glu Ser Gly Val His Gly Ala | | | | | |
| | 1100 | | 1105 | | 1110 |
| Leu Leu Ala Leu Asp Glu Thr Phe Asp Phe Ser Ala Leu Ala Leu | | | | | |
| | 1115 | | 1120 | | 1125 |
| Leu Leu Gln Ile Pro Thr Gln Asn Thr Gln Ala Arg Ala Val Leu | | | | | |
| | 1130 | | 1135 | | 1140 |
| Glu Arg Glu Phe Asn Asn Leu Leu Val Met Gly Thr Asp Arg Arg | | | | | |
| | 1145 | | 1150 | | 1155 |
| Phe Asp Glu Asp Asp Asp Lys Ser Phe Arg Arg Ala Pro Ser Trp | | | | | |
| | 1160 | | 1165 | | 1170 |
| Arg Lys Lys Phe Arg Pro Lys Asp Ile Arg Gly Leu Ala Ala Gly | | | | | |
| | 1175 | | 1180 | | 1185 |
| Ser Ala Glu Thr Leu Pro Ala Asn Phe Arg Val Thr Ser Ser Met | | | | | |
| | 1190 | | 1195 | | 1200 |
| Ser Ser Pro Ser Met Gln Pro Lys Lys Met Gln Met Asp Gly Asn | | | | | |
| | 1205 | | 1210 | | 1215 |
| Val Ser Gly Thr Gln Arg Leu Asp Ser Ala Thr Val Arg Thr Tyr | | | | | |
| | 1220 | | 1225 | | 1230 |
| Ser Cys | | | | | |

<210> 33
 <211> 132
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7509919CD1

<400> 33
 Met Leu Gly Leu Arg Pro Pro Leu Leu Ala Leu Val Gly Leu Leu
 1 5 10 15
 Ser Leu Gly Cys Val Leu Ser Gln Glu Cys Thr Lys Phe Lys Val
 20 25 30
 Ser Ser Cys Arg Glu Cys Ile Glu Ser Gly Pro Gly Cys Thr Trp

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | 35 | | | | | 40 | | | | | 45 |
| Cys | Gln | Lys | Leu | Ala | Arg | Pro | Thr | Leu | Ser | Arg | Leu | Leu | Gly | Leu |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Leu | Glu | Ser | Gly | Gln | Val | Leu | Ser | Phe | Leu | Cys | Leu | Ser | Cys | Pro |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Trp | Ala | Phe | Leu | Arg | Lys | Ser | Gly | Glu | Gln | Leu | Cys | Ser | Met | Thr |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Leu | Gly | Gly | Leu | Val | Trp | Ser | Phe | Leu | Cys | Leu | Asp | Ser | Gly | Ser |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Gly | Phe | Gly | Ser | Ser | Ser | Pro | Arg | His | Arg | Gly | Glu | Gln | Ser | Pro |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Glu | Glu | Lys | Pro | Gly | Gly | Ser | Arg | Leu | Ser | Ser | Asp | | | |
| | | | | 125 | | | | | 130 | | | | | |

<210> 34

<211> 886

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510758CD1

<400> 34

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Glu | Val | Thr | Cys | Leu | Leu | Leu | Leu | Ala | Leu | Ile | Pro | Phe | His |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Cys | Arg | Gly | Gln | Gly | Val | Tyr | Ala | Pro | Ala | Gln | Ala | Gln | Ile | Val |
| | | | | 20 | | | | | 25 | | | | | 30 |
| His | Ala | Gly | Gln | Ala | Cys | Val | Val | Lys | Glu | Asp | Asn | Ile | Ser | Glu |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Arg | Val | Tyr | Thr | Ile | Arg | Glu | Gly | Asp | Thr | Leu | Met | Leu | Gln | Cys |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Leu | Val | Thr | Gly | His | Pro | Arg | Pro | Gln | Val | Arg | Trp | Thr | Lys | Thr |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Ala | Gly | Ser | Ala | Ser | Asp | Lys | Phe | Gln | Glu | Thr | Ser | Val | Phe | Asn |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Glu | Thr | Leu | Arg | Ile | Glu | Arg | Ile | Ala | Arg | Thr | Gln | Gly | Gly | Arg |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Tyr | Tyr | Cys | Lys | Ala | Glu | Asn | Gly | Val | Gly | Val | Pro | Ala | Ile | Lys |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Ser | Ile | Arg | Val | Asp | Val | Gln | Tyr | Leu | Asp | Glu | Pro | Met | Leu | Thr |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Val | His | Gln | Thr | Val | Ser | Asp | Val | Arg | Gly | Asn | Phe | Tyr | Gln | Glu |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Lys | Thr | Val | Phe | Leu | Arg | Cys | Thr | Val | Asn | Ser | Asn | Pro | Pro | Ala |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Arg | Phe | Ile | Trp | Lys | Arg | Gly | Ser | Asp | Thr | Leu | Ser | His | Ser | Gln |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Asp | Asn | Gly | Val | Asp | Ile | Tyr | Glu | Pro | Leu | Tyr | Thr | Gln | Gly | Glu |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Thr | Lys | Val | Leu | Lys | Leu | Lys | Asn | Leu | Arg | Pro | Gln | Asp | Tyr | Ala |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Ser | Tyr | Thr | Cys | Gln | Val | Ser | Val | Arg | Asn | Val | Cys | Gly | Ile | Pro |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Asp | Lys | Ala | Ile | Thr | Phe | Arg | Leu | Thr | Asn | Thr | Thr | Ala | Pro | Pro |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Ala | Leu | Lys | Leu | Ser | Val | Asn | Glu | Thr | Leu | Val | Val | Asn | Pro | Gly |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Glu | Asn | Val | Thr | Val | Gln | Cys | Leu | Leu | Thr | Gly | Gly | Asp | Pro | Leu |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Pro | Gln | Leu | Gln | Trp | Ser | His | Gly | Pro | Gly | Pro | Leu | Pro | Leu | Gly |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Ala | Leu | Ala | Gln | Gly | Gly | Thr | Leu | Ser | Ile | Pro | Ser | Val | Gln | Ala |

| | | | | | |
|-----------------|-----|---------------------|-----|---------------------|-----|
| Arg Asp Ser Gly | 290 | Tyr Tyr Asn Cys Thr | 295 | Thr Asn Asn Val | 300 |
| Asn Pro Ala Lys | 305 | Lys Thr Val Asn Leu | 310 | Val Arg Ser Met | 315 |
| Asn Ala Thr Phe | 320 | Gln Ile Thr Pro Asp | 325 | Ile Lys Glu Ser | 330 |
| Asn Ile Gln Leu | 335 | Gly Gln Asp Leu Lys | 340 | Ser Cys His Val | 345 |
| Ala Val Pro Gln | 350 | Glu Lys Val Thr Tyr | 355 | Gln Trp Phe Lys Asn | 360 |
| Lys Pro Ala Arg | 365 | Met Ser Lys Arg Leu | 370 | Val Thr Arg Asn | 375 |
| Pro Glu Leu Pro | 380 | Ala Val Thr Ser Ser | 385 | Leu Glu Leu Ile Asp | 390 |
| His Phe Ser Asp | 395 | Ala Val Thr Ser Ser | 400 | Glu Leu Ile Asp | 405 |
| Gly Ala Pro Val | 410 | Tyr Gly Thr Tyr Leu | 415 | Cys Met Ala Ser Phe | 420 |
| Glu Thr Val Pro | 425 | Pro Asp Leu Ser Val | 430 | Glu Val Asn Ile Ser | 435 |
| Val Thr Val Arg | 440 | Pro Thr Ile Ser Val | 445 | Pro Lys Gly Arg Ala | 450 |
| Arg Gly Lys Pro | 455 | Glu Gly Ser Pro Ala | 460 | Glu Leu Gln Cys Glu | 465 |
| Glu Ala Ala Leu | 470 | Arg Pro Pro Val Leu | 475 | Trp Ser Arg Val Asp | 480 |
| Asp Gly Lys Leu | 485 | Leu Pro Ser Gly Leu | 490 | Pro Leu Glu Glu Thr | 495 |
| Thr Tyr Arg Cys | 500 | Arg Leu Glu Arg Val | 505 | Ser Arg Asp Met Ser | 510 |
| Pro Arg Glu Ala | 515 | Gln Thr Ala Arg Tyr | 520 | Asn Gly Phe Asn Val | 525 |
| Val Glu Pro Ser | 530 | Gln Val Gln Leu Asn | 535 | Val Gln Phe Pro Pro | 540 |
| Val Leu Leu Arg | 545 | Ser Gln Asp Val Arg | 550 | Gln Ala Leu Gly Arg | 555 |
| Ala Ser Ala Val | 560 | Cys Ser Leu Leu Arg | 565 | Gly Ser Pro Gln Arg | 570 |
| Pro Val Val Pro | 575 | Trp Arg Phe Lys Gly | 580 | Gln Leu Leu Pro Pro | 585 |
| Arg Leu Asp Ala | 590 | Ala Ala Ala Glu Ala | 595 | Pro Asp His Ala Glu | 600 |
| Ser Val Ser Asn | 605 | Val Thr Arg Asp Ser | 610 | Ser Gly Ser Tyr Glu | 615 |
| Ser Ala Lys Ala | 620 | Asp Val Gly Ser Ala | 625 | Ala Cys Leu Phe Gln | 630 |
| Pro Thr Arg Ser | 635 | Tyr Ser Pro Glu Phe | 640 | Phe Asp Thr Pro | 645 |
| Gln Trp Thr Gln | 650 | His Lys Leu Ser Lys | 655 | Asn Tyr Ser Tyr Val | 660 |
| Tyr Arg Leu Ser | 665 | Arg Glu Pro Asp Ala | 670 | Val Asp Pro Val Leu | 675 |
| Lys Ala Ile Pro | 680 | Ile Arg Gln Leu Asn | 685 | Gln His Asn Ala Val | 690 |
| Tyr Ile Leu Thr | 695 | Val Arg Arg Val Glu | 700 | Lys Gly Gln Leu Leu | 705 |
| Leu Thr Pro Tyr | 710 | Asp Leu Arg Val Pro | 715 | His Ser Tyr Glu Val | 720 |
| Ile Ile His Tyr | 725 | Thr Thr Phe Gly Ala | 730 | Gly Asp Met Ala Ser | 735 |
| Asn Thr Cys His | 740 | Thr Glu Pro Ile Asn | 745 | Ser Pro Asn Leu Ser | 750 |
| | 755 | Phe Glu Asp Glu Lys | 760 | Ile Cys Gly Tyr Thr | 765 |

```

Asp Leu Thr Asp Asn Phe Asp Trp Thr Arg Gln Asn Ala Leu Thr
770 775 780
Gln Asn Pro Lys Arg Ser Pro Asn Thr Gly Pro Pro Thr Asp Ile
785 790 795
Ser Gly Thr Pro Glu Gly Tyr Tyr Met Phe Ile Glu Thr Ser Arg
800 805 810
Pro Arg Glu Leu Gly Asp Arg Ala Arg Leu Val Ser Pro Leu Tyr
815 820 825
Asn Ala Ser Ala Lys Phe Tyr Cys Val Ser Phe Phe Tyr His Met
830 835 840
Tyr Gly Lys His Ile Ala Arg Ala Lys Leu His Phe Lys Lys Lys
845 850 855
Lys Gly Lys Arg Lys Asp Ser Ser Arg Met Arg Trp Lys Trp Arg
860 865 870
Gly Trp Arg Arg Gln Leu Arg Gly Cys Lys Tyr Gly Val Thr Ala
875 880 885
Arg

```

```

<210> 35
<211> 859
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7510063CD1

```

```

<400> 35
Met Leu Gln Thr Lys Asp Leu Ile Trp Thr Leu Phe Phe Leu Gly
1 5 10 15
Thr Ala Val Ser Leu Gln Val Asp Ile Val Pro Ser Gln Gly Glu
20 25 30
Ile Ser Val Gly Glu Ser Lys Phe Phe Leu Cys Gln Val Ala Gly
35 40 45
Asp Ala Lys Asp Lys Asp Ile Ser Trp Phe Ser Pro Asn Gly Glu
50 55 60
Lys Leu Thr Pro Asn Gln Gln Arg Ile Ser Val Val Trp Asn Asp
65 70 75
Asp Ser Ser Ser Thr Leu Thr Ile Tyr Asn Ala Asn Ile Asp Asp
80 85 90
Ala Gly Ile Tyr Lys Cys Val Val Thr Gly Glu Asp Gly Ser Glu
95 100 105
Ser Glu Ala Thr Val Asn Val Lys Ile Phe Gln Lys Leu Met Phe
110 115 120
Lys Asn Ala Pro Thr Pro Gln Glu Phe Arg Glu Gly Glu Asp Ala
125 130 135
Val Ile Val Cys Asp Val Val Ser Ser Leu Pro Pro Thr Ile Ile
140 145 150
Trp Lys His Lys Gly Arg Asp Val Ile Leu Lys Lys Asp Val Arg
155 160 165
Phe Ile Val Leu Ser Asn Asn Tyr Leu Gln Ile Arg Gly Ile Lys
170 175 180
Lys Thr Asp Glu Gly Thr Tyr Arg Cys Glu Gly Arg Ile Leu Ala
185 190 195
Arg Gly Glu Ile Asn Phe Lys Asp Ile Gln Val Ile Val Asn Val
200 205 210
Pro Pro Thr Ile Gln Ala Arg Gln Asn Ile Val Asn Ala Thr Ala
215 220 225
Asn Leu Gly Gln Ser Val Thr Leu Val Cys Asp Ala Glu Arg Phe
230 235 240
Pro Glu Pro Thr Met Ser Trp Thr Lys Asp Gly Glu Gln Ile Glu
245 250 255

```

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Glu | Glu | Asp | Asp | Glu | Lys | Tyr | Ile | Phe | Ser | Asp | Asp | Ser | Ser |
| | | | | 260 | | | | | 265 | | | | | 270 |
| Gln | Leu | Thr | Ile | Lys | Lys | Val | Asp | Lys | Asn | Asp | Glu | Ala | Glu | Tyr |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Ile | Cys | Ile | Ala | Glu | Asn | Lys | Ala | Gly | Glu | Gln | Asp | Ala | Thr | Ile |
| | | | | 290 | | | | | 295 | | | | | 300 |
| His | Leu | Lys | Val | Phe | Ala | Lys | Pro | Lys | Ile | Thr | Tyr | Val | Glu | Asn |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Gln | Thr | Ala | Met | Glu | Leu | Glu | Glu | Gln | Val | Thr | Leu | Thr | Cys | Glu |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Ala | Ser | Gly | Asp | Pro | Ile | Pro | Ser | Ile | Thr | Trp | Arg | Thr | Ser | Thr |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Arg | Asn | Ile | Ser | Ser | Glu | Glu | Lys | Ala | Ser | Trp | Thr | Arg | Pro | Glu |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Lys | Gln | Glu | Thr | Leu | Asp | Gly | His | Met | Val | Val | Arg | Ser | His | Ala |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Arg | Val | Ser | Ser | Leu | Thr | Leu | Lys | Ser | Ile | Gln | Tyr | Thr | Asp | Ala |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Gly | Glu | Tyr | Ile | Cys | Thr | Ala | Ser | Asn | Thr | Ile | Gly | Gln | Asp | Ser |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Gln | Ser | Met | Tyr | Leu | Glu | Val | Gln | Tyr | Ala | Pro | Lys | Leu | Gln | Gly |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Pro | Val | Ala | Val | Tyr | Thr | Trp | Glu | Gly | Asn | Gln | Val | Asn | Ile | Thr |
| | | | | 425 | | | | | 430 | | | | | 435 |
| Cys | Glu | Val | Phe | Ala | Tyr | Pro | Ser | Ala | Thr | Ile | Ser | Trp | Phe | Arg |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Asp | Gly | Gln | Leu | Leu | Pro | Ser | Ser | Asn | Tyr | Ser | Asn | Ile | Lys | Ile |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Tyr | Asn | Thr | Pro | Ser | Ala | Ser | Tyr | Leu | Glu | Val | Thr | Pro | Asp | Ser |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Glu | Asn | Asp | Phe | Gly | Asn | Tyr | Asn | Cys | Thr | Ala | Val | Asn | Arg | Ile |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Gly | Gln | Glu | Ser | Leu | Glu | Phe | Ile | Leu | Val | Gln | Ala | Asp | Thr | Pro |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Ser | Ser | Pro | Ser | Ile | Asp | Gln | Val | Glu | Pro | Tyr | Ser | Ser | Thr | Ala |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Gln | Val | Gln | Phe | Asp | Glu | Pro | Glu | Ala | Thr | Gly | Gly | Val | Pro | Ile |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Leu | Lys | Tyr | Lys | Ala | Glu | Trp | Arg | Ala | Val | Gly | Glu | Glu | Val | Trp |
| | | | | 545 | | | | | 550 | | | | | 555 |
| His | Ser | Lys | Trp | Tyr | Asp | Ala | Lys | Glu | Ala | Ser | Met | Glu | Gly | Ile |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Val | Thr | Ile | Val | Gly | Leu | Lys | Pro | Glu | Thr | Thr | Tyr | Ala | Val | Arg |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Leu | Ala | Ala | Leu | Asn | Gly | Lys | Gly | Leu | Gly | Glu | Ile | Ser | Ala | Ala |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Ser | Glu | Phe | Lys | Thr | Gln | Pro | Val | Gln | Gly | Glu | Pro | Ser | Ala | Pro |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Lys | Leu | Glu | Gly | Gln | Met | Gly | Glu | Asp | Gly | Asn | Ser | Ile | Lys | Val |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Asn | Leu | Ile | Lys | Gln | Asp | Asp | Gly | Gly | Ser | Pro | Ile | Arg | His | Tyr |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Leu | Val | Arg | Tyr | Arg | Ala | Lys | Leu | Ser | Ser | Glu | Trp | Lys | Pro | Glu |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Ile | Arg | Leu | Pro | Ser | Gly | Ser | Asp | His | Val | Met | Leu | Lys | Ser | Leu |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Asp | Trp | Asn | Ala | Glu | Tyr | Glu | Val | Tyr | Val | Val | Ala | Glu | Asn | Gln |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Gln | Gly | Lys | Ser | Lys | Ala | Ala | His | Phe | Val | Phe | Arg | Thr | Ser | Ala |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Gln | Pro | Thr | Ala | Ile | Pro | Ala | Asn | Gly | Ser | Pro | Thr | Ser | Gly | Leu |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Ser | Thr | Gly | Ala | Ile | Val | Gly | Ile | Leu | Ile | Val | Ile | Phe | Val | Leu |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | | | 725 | | | | | 730 | | | | 735 | |
| Leu | Leu | Val | Val | Val | Asp | Ile | Thr | Cys | Tyr | Phe | Leu | Asn | Lys | Cys |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Gly | Leu | Phe | Met | Cys | Ile | Ala | Val | Asn | Leu | Cys | Gly | Lys | Ala | Gly |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Pro | Gly | Ala | Lys | Gly | Lys | Asp | Met | Glu | Glu | Gly | Lys | Ala | Ala | Phe |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Ser | Lys | Asp | Glu | Ser | Lys | Glu | Pro | Ile | Val | Glu | Val | Arg | Thr | Glu |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Glu | Glu | Arg | Thr | Pro | Asn | His | Asp | Gly | Gly | Lys | His | Thr | Glu | Pro |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Asn | Glu | Thr | Thr | Pro | Leu | Thr | Glu | Pro | Glu | Lys | Gly | Pro | Val | Glu |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Ala | Lys | Pro | Glu | Cys | Gln | Glu | Thr | Glu | Thr | Lys | Pro | Ala | Pro | Ala |
| | | | | 830 | | | | | 835 | | | | | 840 |
| Glu | Val | Lys | Thr | Val | Pro | Asn | Asp | Ala | Thr | Gln | Thr | Lys | Glu | Asn |
| | | | | 845 | | | | | 850 | | | | | 855 |
| Glu | Ser | Lys | Ala | | | | | | | | | | | |

<210> 36

<211> 195

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510135CD1

<400> 36

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Ala | Phe | Pro | Pro | Arg | Arg | Arg | Leu | Arg | Leu | Gly | Pro | Arg | Gly |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Leu | Pro | Leu | Leu | Leu | Ser | Gly | Leu | Leu | Pro | Leu | Cys | Arg | Ala | |
| | | | | 20 | | | | | 25 | | | | | 30 |
| Phe | Asn | Leu | Asp | Val | Asp | Ser | Pro | Ala | Glu | Tyr | Ser | Gly | Pro | Glu |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Gly | Ser | Tyr | Phe | Gly | Phe | Ala | Val | Asp | Phe | Phe | Val | Pro | Ser | Ala |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Ser | Ser | Arg | Met | Phe | Leu | Leu | Val | Gly | Ala | Pro | Lys | Ala | Asn | Thr |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Thr | Gln | Pro | Gly | Ile | Val | Glu | Gly | Gly | Gln | Val | Leu | Lys | Cys | Asp |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Trp | Ser | Ser | Thr | Arg | Arg | Cys | Gln | Pro | Ile | Glu | Phe | Asp | Ala | Thr |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Gly | Asn | Arg | Asp | Tyr | Ala | Lys | Asp | Asp | Pro | Leu | Glu | Phe | Lys | Ser |
| | | | | 110 | | | | | 115 | | | | | 120 |
| His | Gln | Trp | Phe | Gly | Ala | Ser | Val | Arg | Ser | Lys | Gln | Asp | Lys | Ile |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Leu | Ala | Cys | Ala | Pro | Leu | Tyr | His | Trp | Arg | Thr | Glu | Met | Lys | Gln |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Glu | Arg | Glu | Pro | Val | Gly | Thr | Cys | Phe | Leu | Gln | Asp | Gly | Thr | Lys |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Thr | Val | Glu | Tyr | Ala | Pro | Cys | Arg | Ser | Gln | Asp | Ile | Asp | Ala | Asp |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Gly | Gln | Gly | Phe | Cys | Gln | Gly | Gly | Phe | Ser | Ile | Asp | Phe | Thr | Lys |
| | | | | 185 | | | | | 190 | | | | | 195 |

<210> 37

<211> 2110

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7505011CD1

<400> 37

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Gly | Ala | Met | Thr | Gln | Leu | Leu | Ala | Gly | Val | Phe | Leu | Ala | Phe |
| 1 | | | | 5 | | | | | 10 | | | | | 15 |
| Leu | Ala | Leu | Ala | Thr | Glu | Gly | Gly | Val | Leu | Lys | Lys | Val | Ile | Arg |
| | | | | 20 | | | | | 25 | | | | | 30 |
| His | Lys | Arg | Gln | Ser | Gly | Val | Asn | Ala | Thr | Leu | Pro | Glu | Glu | Asn |
| | | | | 35 | | | | | 40 | | | | | 45 |
| Gln | Pro | Val | Val | Phe | Asn | His | Val | Tyr | Asn | Ile | Lys | Leu | Pro | Val |
| | | | | 50 | | | | | 55 | | | | | 60 |
| Gly | Ser | Gln | Cys | Ser | Val | Asp | Leu | Glu | Ser | Ala | Ser | Gly | Glu | Lys |
| | | | | 65 | | | | | 70 | | | | | 75 |
| Asp | Leu | Ala | Pro | Pro | Ser | Glu | Pro | Ser | Glu | Ser | Phe | Gln | Glu | His |
| | | | | 80 | | | | | 85 | | | | | 90 |
| Thr | Val | Asp | Gly | Glu | Asn | Gln | Ile | Val | Phe | Thr | His | Arg | Ile | Asn |
| | | | | 95 | | | | | 100 | | | | | 105 |
| Ile | Pro | Arg | Arg | Ala | Cys | Gly | Cys | Ala | Ala | Ala | Pro | Asp | Val | Lys |
| | | | | 110 | | | | | 115 | | | | | 120 |
| Glu | Leu | Leu | Ser | Arg | Leu | Glu | Glu | Leu | Glu | Asn | Leu | Val | Ser | Ser |
| | | | | 125 | | | | | 130 | | | | | 135 |
| Leu | Arg | Glu | Gln | Cys | Thr | Ala | Gly | Ala | Gly | Cys | Cys | Leu | Gln | Pro |
| | | | | 140 | | | | | 145 | | | | | 150 |
| Ala | Thr | Gly | Arg | Leu | Asp | Thr | Arg | Pro | Phe | Cys | Ser | Gly | Arg | Gly |
| | | | | 155 | | | | | 160 | | | | | 165 |
| Asn | Phe | Ser | Thr | Glu | Gly | Cys | Gly | Cys | Val | Cys | Glu | Pro | Gly | Trp |
| | | | | 170 | | | | | 175 | | | | | 180 |
| Lys | Gly | Pro | Asn | Cys | Ser | Glu | Pro | Glu | Cys | Pro | Gly | Asn | Cys | His |
| | | | | 185 | | | | | 190 | | | | | 195 |
| Leu | Arg | Gly | Arg | Cys | Ile | Asp | Gly | Gln | Cys | Ile | Cys | Asp | Asp | Gly |
| | | | | 200 | | | | | 205 | | | | | 210 |
| Phe | Thr | Gly | Glu | Asp | Cys | Ser | Gln | Leu | Ala | Cys | Pro | Ser | Asp | Cys |
| | | | | 215 | | | | | 220 | | | | | 225 |
| Asn | Asp | Gln | Gly | Lys | Cys | Val | Asn | Gly | Val | Cys | Ile | Cys | Phe | Glu |
| | | | | 230 | | | | | 235 | | | | | 240 |
| Gly | Tyr | Ala | Gly | Ala | Asp | Cys | Ser | Arg | Glu | Ile | Cys | Pro | Val | Pro |
| | | | | 245 | | | | | 250 | | | | | 255 |
| Cys | Ser | Glu | Glu | His | Gly | Thr | Cys | Val | Asp | Gly | Leu | Cys | Val | Cys |
| | | | | 260 | | | | | 265 | | | | | 270 |
| His | Asp | Gly | Phe | Ala | Gly | Asp | Asp | Cys | Asn | Lys | Pro | Leu | Cys | Leu |
| | | | | 275 | | | | | 280 | | | | | 285 |
| Asn | Asn | Cys | Tyr | Asn | Arg | Gly | Arg | Cys | Val | Glu | Asn | Glu | Cys | Val |
| | | | | 290 | | | | | 295 | | | | | 300 |
| Cys | Asp | Glu | Gly | Phe | Thr | Gly | Glu | Asp | Cys | Ser | Glu | Leu | Ile | Cys |
| | | | | 305 | | | | | 310 | | | | | 315 |
| Pro | Asn | Asp | Cys | Phe | Asp | Arg | Gly | Arg | Cys | Ile | Asn | Gly | Thr | Cys |
| | | | | 320 | | | | | 325 | | | | | 330 |
| Tyr | Cys | Glu | Glu | Gly | Phe | Thr | Gly | Glu | Asp | Cys | Gly | Lys | Pro | Thr |
| | | | | 335 | | | | | 340 | | | | | 345 |
| Cys | Pro | His | Ala | Cys | His | Thr | Gln | Gly | Arg | Cys | Glu | Glu | Gly | Gln |
| | | | | 350 | | | | | 355 | | | | | 360 |
| Cys | Val | Cys | Asp | Glu | Gly | Phe | Ala | Gly | Val | Asp | Cys | Ser | Glu | Lys |
| | | | | 365 | | | | | 370 | | | | | 375 |
| Arg | Cys | Pro | Ala | Asp | Cys | His | Asn | Arg | Gly | Arg | Cys | Val | Asp | Gly |
| | | | | 380 | | | | | 385 | | | | | 390 |
| Arg | Cys | Glu | Cys | Asp | Asp | Gly | Phe | Thr | Gly | Ala | Asp | Cys | Gly | Glu |
| | | | | 395 | | | | | 400 | | | | | 405 |
| Leu | Lys | Cys | Pro | Asn | Gly | Cys | Ser | Gly | His | Gly | Arg | Cys | Val | Asn |
| | | | | 410 | | | | | 415 | | | | | 420 |
| Gly | Gln | Cys | Val | Cys | Asp | Glu | Gly | Tyr | Thr | Gly | Glu | Asp | Cys | Ser |
| | | | | 425 | | | | | 430 | | | | | 435 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Leu | Arg | Cys | Pro | Asn | Asp | Cys | His | Ser | Arg | Gly | Arg | Cys | Val |
| | | | | 440 | | | | | 445 | | | | | 450 |
| Glu | Gly | Lys | Cys | Val | Cys | Glu | Gln | Gly | Phe | Lys | Gly | Tyr | Asp | Cys |
| | | | | 455 | | | | | 460 | | | | | 465 |
| Ser | Asp | Met | Ser | Cys | Pro | Asn | Asp | Cys | His | Gln | His | Gly | Arg | Cys |
| | | | | 470 | | | | | 475 | | | | | 480 |
| Val | Asn | Gly | Met | Cys | Val | Cys | Asp | Asp | Gly | Tyr | Thr | Gly | Glu | Asp |
| | | | | 485 | | | | | 490 | | | | | 495 |
| Cys | Arg | Asp | Arg | Gln | Cys | Pro | Arg | Asp | Cys | Ser | Asn | Arg | Gly | Leu |
| | | | | 500 | | | | | 505 | | | | | 510 |
| Cys | Val | Asp | Gly | Gln | Cys | Val | Cys | Glu | Asp | Gly | Phe | Thr | Gly | Pro |
| | | | | 515 | | | | | 520 | | | | | 525 |
| Asp | Cys | Ala | Glu | Leu | Ser | Cys | Pro | Asn | Asp | Cys | His | Gly | Arg | Gly |
| | | | | 530 | | | | | 535 | | | | | 540 |
| Arg | Cys | Val | Asn | Gly | Gln | Cys | Val | Cys | His | Glu | Gly | Phe | Met | Gly |
| | | | | 545 | | | | | 550 | | | | | 555 |
| Lys | Asp | Cys | Lys | Glu | Gln | Arg | Cys | Pro | Ser | Asp | Cys | His | Gly | Gln |
| | | | | 560 | | | | | 565 | | | | | 570 |
| Gly | Arg | Cys | Val | Asp | Gly | Gln | Cys | Ile | Cys | His | Glu | Gly | Phe | Thr |
| | | | | 575 | | | | | 580 | | | | | 585 |
| Gly | Leu | Asp | Cys | Gly | Gln | His | Ser | Cys | Pro | Ser | Asp | Cys | Asn | Asn |
| | | | | 590 | | | | | 595 | | | | | 600 |
| Leu | Gly | Gln | Cys | Val | Ser | Gly | Arg | Cys | Ile | Cys | Asn | Glu | Gly | Tyr |
| | | | | 605 | | | | | 610 | | | | | 615 |
| Ser | Gly | Glu | Asp | Cys | Ser | Glu | Val | Ser | Pro | Pro | Lys | Asp | Leu | Val |
| | | | | 620 | | | | | 625 | | | | | 630 |
| Val | Thr | Glu | Val | Thr | Glu | Glu | Thr | Val | Asn | Leu | Ala | Trp | Asp | Asn |
| | | | | 635 | | | | | 640 | | | | | 645 |
| Glu | Met | Arg | Val | Thr | Glu | Tyr | Leu | Val | Val | Tyr | Thr | Pro | Thr | His |
| | | | | 650 | | | | | 655 | | | | | 660 |
| Glu | Gly | Gly | Leu | Glu | Met | Gln | Phe | Arg | Val | Pro | Gly | Asp | Gln | Thr |
| | | | | 665 | | | | | 670 | | | | | 675 |
| Ser | Thr | Ile | Ile | Gln | Glu | Leu | Glu | Pro | Gly | Val | Glu | Tyr | Phe | Ile |
| | | | | 680 | | | | | 685 | | | | | 690 |
| Arg | Val | Phe | Ala | Ile | Leu | Glu | Asn | Lys | Lys | Ser | Ile | Pro | Val | Ser |
| | | | | 695 | | | | | 700 | | | | | 705 |
| Ala | Arg | Val | Ala | Thr | Tyr | Leu | Pro | Ala | Pro | Glu | Gly | Leu | Lys | Phe |
| | | | | 710 | | | | | 715 | | | | | 720 |
| Lys | Ser | Ile | Lys | Glu | Thr | Ser | Val | Glu | Val | Glu | Trp | Asp | Pro | Leu |
| | | | | 725 | | | | | 730 | | | | | 735 |
| Asp | Ile | Ala | Phe | Glu | Thr | Trp | Glu | Ile | Ile | Phe | Arg | Asn | Met | Asn |
| | | | | 740 | | | | | 745 | | | | | 750 |
| Lys | Glu | Asp | Glu | Gly | Glu | Ile | Thr | Lys | Ser | Leu | Arg | Arg | Pro | Glu |
| | | | | 755 | | | | | 760 | | | | | 765 |
| Thr | Ser | Tyr | Arg | Gln | Thr | Gly | Leu | Ala | Pro | Gly | Gln | Glu | Tyr | Glu |
| | | | | 770 | | | | | 775 | | | | | 780 |
| Ile | Ser | Leu | His | Ile | Val | Lys | Asn | Asn | Thr | Arg | Gly | Pro | Gly | Leu |
| | | | | 785 | | | | | 790 | | | | | 795 |
| Lys | Arg | Val | Thr | Thr | Thr | Arg | Leu | Asp | Ala | Pro | Ser | Gln | Ile | Glu |
| | | | | 800 | | | | | 805 | | | | | 810 |
| Val | Lys | Asp | Val | Thr | Asp | Thr | Thr | Ala | Leu | Ile | Thr | Trp | Phe | Lys |
| | | | | 815 | | | | | 820 | | | | | 825 |
| Pro | Leu | Ala | Glu | Ile | Asp | Gly | Ile | Glu | Leu | Thr | Tyr | Gly | Ile | Lys |
| | | | | 830 | | | | | 835 | | | | | 840 |
| Asp | Val | Pro | Gly | Asp | Arg | Thr | Thr | Ile | Asp | Leu | Thr | Glu | Asp | Glu |
| | | | | 845 | | | | | 850 | | | | | 855 |
| Asn | Gln | Tyr | Ser | Ile | Gly | Asn | Leu | Lys | Pro | Asp | Thr | Glu | Tyr | Glu |
| | | | | 860 | | | | | 865 | | | | | 870 |
| Val | Ser | Leu | Ile | Ser | Arg | Arg | Gly | Asp | Met | Ser | Ser | Asn | Pro | Ala |
| | | | | 875 | | | | | 880 | | | | | 885 |
| Lys | Glu | Thr | Phe | Thr | Thr | Gly | Leu | Asp | Ala | Pro | Arg | Asn | Leu | Arg |
| | | | | 890 | | | | | 895 | | | | | 900 |
| Arg | Val | Ser | Gln | Thr | Asp | Asn | Ser | Ile | Thr | Leu | Glu | Trp | Arg | Asn |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|
| | | | | 905 | | | | | 910 | | | | | 915 |
| Gly | Lys | Ala | Ala | Ile | Asp | Ser | Tyr | Arg | Ile | Lys | Tyr | Ala | Pro | Ile |
| | | | | 920 | | | | | 925 | | | | | 930 |
| Ser | Gly | Gly | Asp | His | Ala | Glu | Val | Asp | Val | Pro | Lys | Ser | Gln | Gln |
| | | | | 935 | | | | | 940 | | | | | 945 |
| Ala | Thr | Thr | Lys | Thr | Thr | Leu | Thr | Gly | Leu | Arg | Pro | Gly | Thr | Glu |
| | | | | 950 | | | | | 955 | | | | | 960 |
| Tyr | Gly | Ile | Gly | Val | Ser | Ala | Val | Lys | Glu | Asp | Lys | Glu | Ser | Asn |
| | | | | 965 | | | | | 970 | | | | | 975 |
| Pro | Ala | Thr | Ile | Asn | Ala | Ala | Thr | Glu | Leu | Asp | Thr | Pro | Lys | Asp |
| | | | | 980 | | | | | 985 | | | | | 990 |
| Leu | Gln | Val | Ser | Glu | Thr | Ala | Glu | Thr | Ser | Leu | Thr | Leu | Leu | Trp |
| | | | | 995 | | | | | 1000 | | | | | 1005 |
| Lys | Thr | Pro | Leu | Ala | Lys | Phe | Asp | Arg | Tyr | Arg | Leu | Asn | Tyr | Ser |
| | | | | 1010 | | | | | 1015 | | | | | 1020 |
| Leu | Pro | Thr | Gly | Gln | Trp | Val | Gly | Val | Gln | Leu | Pro | Arg | Asn | Thr |
| | | | | 1025 | | | | | 1030 | | | | | 1035 |
| Thr | Ser | Tyr | Val | Leu | Arg | Gly | Leu | Glu | Pro | Gly | Gln | Glu | Tyr | Asn |
| | | | | 1040 | | | | | 1045 | | | | | 1050 |
| Val | Leu | Leu | Thr | Ala | Glu | Lys | Gly | Arg | His | Lys | Ser | Lys | Pro | Ala |
| | | | | 1055 | | | | | 1060 | | | | | 1065 |
| Arg | Val | Lys | Ala | Ser | Thr | Glu | Arg | Ala | Pro | Glu | Leu | Glu | Asn | Leu |
| | | | | 1070 | | | | | 1075 | | | | | 1080 |
| Thr | Val | Thr | Glu | Val | Gly | Trp | Asp | Gly | Leu | Arg | Leu | Asn | Trp | Thr |
| | | | | 1085 | | | | | 1090 | | | | | 1095 |
| Ala | Ala | Asp | Gln | Ala | Tyr | Glu | His | Phe | Ile | Ile | Gln | Val | Gln | Glu |
| | | | | 1100 | | | | | 1105 | | | | | 1110 |
| Ala | Asn | Lys | Val | Glu | Ala | Ala | Arg | Asn | Leu | Thr | Val | Pro | Gly | Ser |
| | | | | 1115 | | | | | 1120 | | | | | 1125 |
| Leu | Arg | Ala | Val | Asp | Ile | Pro | Gly | Leu | Lys | Ala | Ala | Thr | Pro | Tyr |
| | | | | 1130 | | | | | 1135 | | | | | 1140 |
| Thr | Val | Ser | Ile | Tyr | Gly | Ser | Phe | Gln | Gly | Tyr | Arg | Thr | Pro | Val |
| | | | | 1145 | | | | | 1150 | | | | | 1155 |
| Leu | Ser | Ala | Glu | Ala | Ser | Thr | Gly | Glu | Thr | Pro | Asn | Leu | Gly | Glu |
| | | | | 1160 | | | | | 1165 | | | | | 1170 |
| Val | Val | Val | Ala | Glu | Val | Gly | Trp | Asp | Ala | Leu | Lys | Leu | Asn | Trp |
| | | | | 1175 | | | | | 1180 | | | | | 1185 |
| Thr | Ala | Pro | Glu | Gly | Ala | Tyr | Glu | Tyr | Phe | Phe | Ile | Gln | Val | Gln |
| | | | | 1190 | | | | | 1195 | | | | | 1200 |
| Glu | Ala | Asp | Thr | Val | Glu | Ala | Ala | Gln | Asn | Leu | Thr | Val | Pro | Gly |
| | | | | 1205 | | | | | 1210 | | | | | 1215 |
| Gly | Leu | Arg | Ser | Thr | Asp | Leu | Pro | Gly | Leu | Lys | Ala | Ala | Thr | His |
| | | | | 1220 | | | | | 1225 | | | | | 1230 |
| Tyr | Thr | Ile | Thr | Ile | Arg | Gly | Val | Thr | Gln | Asp | Phe | Ser | Thr | Thr |
| | | | | 1235 | | | | | 1240 | | | | | 1245 |
| Pro | Leu | Ser | Val | Glu | Val | Leu | Thr | Glu | Glu | Val | Pro | Asp | Met | Gly |
| | | | | 1250 | | | | | 1255 | | | | | 1260 |
| Asn | Leu | Thr | Val | Thr | Glu | Val | Ser | Trp | Asp | Ala | Leu | Arg | Leu | Asn |
| | | | | 1265 | | | | | 1270 | | | | | 1275 |
| Trp | Thr | Thr | Pro | Asp | Gly | Thr | Tyr | Asp | Gln | Phe | Thr | Ile | Gln | Val |
| | | | | 1280 | | | | | 1285 | | | | | 1290 |
| Gln | Glu | Ala | Asp | Gln | Val | Glu | Glu | Ala | His | Asn | Leu | Thr | Val | Pro |
| | | | | 1295 | | | | | 1300 | | | | | 1305 |
| Gly | Ser | Leu | Arg | Ser | Met | Glu | Ile | Pro | Gly | Leu | Arg | Ala | Gly | Thr |
| | | | | 1310 | | | | | 1315 | | | | | 1320 |
| Pro | Tyr | Thr | Val | Thr | Leu | His | Gly | Glu | Val | Arg | Gly | His | Ser | Thr |
| | | | | 1325 | | | | | 1330 | | | | | 1335 |
| Arg | Pro | Leu | Ala | Val | Glu | Val | Val | Thr | Glu | Asp | Leu | Pro | Gln | Leu |
| | | | | 1340 | | | | | 1345 | | | | | 1350 |
| Gly | Asp | Leu | Ala | Val | Ser | Glu | Val | Gly | Trp | Asp | Gly | Leu | Arg | Leu |
| | | | | 1355 | | | | | 1360 | | | | | 1365 |
| Asn | Trp | Thr | Ala | Ala | Asp | Asn | Ala | Tyr | Glu | His | Phe | Val | Ile | Gln |
| | | | | 1370 | | | | | 1375 | | | | | 1380 |

| | | | | | | | | | | | | | | |
|-----|-----|-----|-----|------|-----|-----|-----|-----|------|-----|-----|-----|-----|------|
| Val | Gln | Glu | Val | Asn | Lys | Val | Glu | Ala | Ala | Gln | Asn | Leu | Thr | Leu |
| | | | | 1385 | | | | | 1390 | | | | | 1395 |
| Pro | Gly | Ser | Leu | Arg | Ala | Val | Asp | Ile | Pro | Gly | Leu | Glu | Ala | Ala |
| | | | | 1400 | | | | | 1405 | | | | | 1410 |
| Thr | Pro | Tyr | Arg | Val | Ser | Ile | Tyr | Gly | Val | Ile | Arg | Gly | Tyr | Arg |
| | | | | 1415 | | | | | 1420 | | | | | 1425 |
| Thr | Pro | Val | Leu | Ser | Ala | Glu | Ala | Ser | Thr | Ala | Lys | Glu | Pro | Glu |
| | | | | 1430 | | | | | 1435 | | | | | 1440 |
| Ile | Gly | Asn | Leu | Asn | Val | Ser | Asp | Ile | Thr | Pro | Glu | Ser | Phe | Asn |
| | | | | 1445 | | | | | 1450 | | | | | 1455 |
| Leu | Ser | Trp | Met | Ala | Thr | Asp | Gly | Ile | Phe | Glu | Thr | Phe | Thr | Ile |
| | | | | 1460 | | | | | 1465 | | | | | 1470 |
| Glu | Ile | Ile | Asp | Ser | Asn | Arg | Leu | Leu | Glu | Thr | Val | Glu | Tyr | Asn |
| | | | | 1475 | | | | | 1480 | | | | | 1485 |
| Ile | Ser | Gly | Ala | Glu | Arg | Thr | Ala | His | Ile | Ser | Gly | Leu | Pro | Pro |
| | | | | 1490 | | | | | 1495 | | | | | 1500 |
| Ser | Thr | Asp | Phe | Ile | Val | Tyr | Leu | Ser | Gly | Leu | Ala | Pro | Ser | Ile |
| | | | | 1505 | | | | | 1510 | | | | | 1515 |
| Arg | Thr | Lys | Thr | Ile | Ser | Ala | Thr | Ala | Thr | Thr | Glu | Ala | Glu | Pro |
| | | | | 1520 | | | | | 1525 | | | | | 1530 |
| Glu | Val | Asp | Asn | Leu | Leu | Val | Ser | Asp | Ala | Thr | Pro | Asp | Gly | Phe |
| | | | | 1535 | | | | | 1540 | | | | | 1545 |
| Arg | Leu | Ser | Trp | Thr | Ala | Asp | Glu | Gly | Val | Phe | Asp | Asn | Phe | Val |
| | | | | 1550 | | | | | 1555 | | | | | 1560 |
| Leu | Lys | Ile | Arg | Asp | Thr | Lys | Lys | Gln | Ser | Glu | Pro | Leu | Glu | Ile |
| | | | | 1565 | | | | | 1570 | | | | | 1575 |
| Thr | Leu | Leu | Ala | Pro | Glu | Arg | Thr | Arg | Asp | Leu | Thr | Gly | Leu | Arg |
| | | | | 1580 | | | | | 1585 | | | | | 1590 |
| Glu | Ala | Thr | Glu | Tyr | Glu | Ile | Glu | Leu | Tyr | Gly | Ile | Ser | Lys | Gly |
| | | | | 1595 | | | | | 1600 | | | | | 1605 |
| Arg | Arg | Ser | Gln | Thr | Val | Ser | Ala | Ile | Ala | Thr | Thr | Ala | Met | Gly |
| | | | | 1610 | | | | | 1615 | | | | | 1620 |
| Ser | Pro | Lys | Glu | Val | Ile | Phe | Ser | Asp | Ile | Thr | Glu | Asn | Ser | Ala |
| | | | | 1625 | | | | | 1630 | | | | | 1635 |
| Thr | Val | Ser | Trp | Arg | Ala | Pro | Thr | Ala | Gln | Val | Glu | Ser | Phe | Arg |
| | | | | 1640 | | | | | 1645 | | | | | 1650 |
| Ile | Thr | Tyr | Val | Pro | Ile | Thr | Gly | Gly | Thr | Pro | Ser | Met | Val | Thr |
| | | | | 1655 | | | | | 1660 | | | | | 1665 |
| Val | Asp | Gly | Thr | Lys | Thr | Gln | Thr | Arg | Leu | Val | Lys | Leu | Ile | Pro |
| | | | | 1670 | | | | | 1675 | | | | | 1680 |
| Gly | Val | Glu | Tyr | Leu | Val | Ser | Ile | Ile | Ala | Met | Lys | Gly | Phe | Glu |
| | | | | 1685 | | | | | 1690 | | | | | 1695 |
| Glu | Ser | Glu | Pro | Val | Ser | Gly | Ser | Phe | Thr | Thr | Ala | Leu | Asp | Gly |
| | | | | 1700 | | | | | 1705 | | | | | 1710 |
| Pro | Ser | Gly | Leu | Val | Thr | Ala | Asn | Ile | Thr | Asp | Ser | Glu | Ala | Leu |
| | | | | 1715 | | | | | 1720 | | | | | 1725 |
| Ala | Arg | Trp | Gln | | | | | | | | | | | |

| | | |
|---|------|------|
| 1850 | 1855 | 1860 |
| Thr Ala Lys Ile Gln Ala Leu Asn Gly Pro Leu Arg Ser Asn Met | | |
| 1865 | 1870 | 1875 |
| Ile Gln Thr Ile Phe Thr Thr Ile Gly Leu Leu Tyr Pro Phe Pro | | |
| 1880 | 1885 | 1890 |
| Lys Asp Cys Ser Gln Ala Met Leu Asn Gly Asp Thr Thr Ser Gly | | |
| 1895 | 1900 | 1905 |
| Leu Tyr Thr Ile Tyr Leu Asn Gly Asp Lys Ala Gln Ala Leu Glu | | |
| 1910 | 1915 | 1920 |
| Val Phe Cys Asp Met Thr Ser Asp Gly Gly Gly Trp Ile Val Phe | | |
| 1925 | 1930 | 1935 |
| Leu Arg Arg Lys Asn Gly Arg Glu Asn Phe Tyr Gln Asn Trp Lys | | |
| 1940 | 1945 | 1950 |
| Ala Tyr Ala Ala Gly Phe Gly Asp Arg Arg Glu Glu Phe Trp Leu | | |
| 1955 | 1960 | 1965 |
| Gly Leu Asp Asn Leu Asn Lys Ile Thr Ala Gln Gly Gln Tyr Glu | | |
| 1970 | 1975 | 1980 |
| Leu Arg Val Asp Leu Arg Asp His Gly Glu Thr Ala Phe Ala Val | | |
| 1985 | 1990 | 1995 |
| Tyr Asp Lys Phe Ser Val Gly Asp Ala Lys Thr Arg Tyr Lys Leu | | |
| 2000 | 2005 | 2010 |
| Lys Val Glu Gly Tyr Ser Gly Thr Ala Gly Asp Ser Met Ala Tyr | | |
| 2015 | 2020 | 2025 |
| His Asn Gly Arg Ser Phe Ser Thr Phe Asp Lys Asp Thr Asp Ser | | |
| 2030 | 2035 | 2040 |
| Ala Ile Thr Asn Cys Ala Leu Ser Tyr Lys Gly Ala Phe Trp Tyr | | |
| 2045 | 2050 | 2055 |
| Arg Asn Cys His Arg Val Asn Leu Met Gly Arg Tyr Gly Asp Asn | | |
| 2060 | 2065 | 2070 |
| Asn His Ser Gln Gly Val Asn Trp Phe His Trp Lys Gly His Glu | | |
| 2075 | 2080 | 2085 |
| His Ser Ile Gln Phe Ala Glu Met Lys Leu Arg Pro Ser Asn Phe | | |
| 2090 | 2095 | 2100 |
| Arg Asn Leu Glu Gly Arg Arg Lys Arg Ala | | |
| 2105 | 2110 | |

<210> 38

<211> 1469

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7504868CB1

<400> 38

| | | | | | | |
|-------------|------------|------------|------------|------------|-------------|------|
| tacccccccc | cccgcccccc | acgccgatgc | aagctggtac | gagctcgcat | cgatagtaac | 60 |
| ggcgcagtgt | gctggaatcg | ccttaaggag | tgggcgcctc | tatttaagcg | gcttccccgc | 120 |
| ggcctcggga | cagaggggac | tgagcatgga | tttcggactg | gccctcctgc | tggcggggct | 180 |
| tctggggctc | ctcctcgcc | tcccggacca | gctgaccgtc | tccccggcag | ccctggtgcc | 240 |
| tggtgacccg | gaggtggcct | gtacggccca | caaagtcacg | cccgtggacc | ccaacgcgct | 300 |
| ctccttctcc | ctgctcgtcg | ggggccagga | actggagggg | gcgcaagccc | tgggcccggga | 360 |
| ggtgcaggag | gaggaggagg | agccccaggg | ggacgaggac | gtgctgttca | gggtgacaga | 420 |
| gcgctggcgg | ctgccgcccc | tggggacccc | tgtcccgcgc | gccctctact | gccaggccac | 480 |
| gatgaggctg | cctggcttgg | agctcagcca | ccgccaggcc | atccccgtcc | tgcacagccc | 540 |
| gacctccccg | gagcctcccc | acaccacctc | cccggagcct | cccaacacca | cctccccgga | 600 |
| gtctccccgac | accacctccc | cggagtctcc | cgacaccacc | tcccaggagc | ctccccgacac | 660 |
| cacctcccag | gagcctcccc | acaccacctc | ccaggagcct | cccgaaccca | cctccccgga | 720 |
| gcctccccgac | aagacctccc | cggagcccg | ccccagcag | ggctccacac | acacccccag | 780 |
| gagcccaggc | tccaccagga | ctcgccgcgc | tgagatctcc | caggctgggc | ccacgcaggg | 840 |
| agaagtgatc | ccaacaggct | cgtccaaacc | tgcggtgac | cagctgccc | cggtctctgtg | 900 |
| gaccagcagt | gcggtgctgg | gactgctgct | cctggccttg | cccacctatc | acctctggaa | 960 |
| acgctgccgg | cacctggctg | aggacgacac | ccaccaccca | gcttctctga | ggcttctgcc | 1020 |

| | | | | | | |
|-------------|-------------|-------------|------------|------------|------------|------|
| ccagggtgtcg | gcctgggctg | gggttaagggg | gaccggccag | gtcgggatca | gccccctctg | 1080 |
| agtgggccagc | ctttcccccct | gtgaaagcaa | aatagccttg | accccttcaa | gttgagaact | 1140 |
| gggtcagggca | aacctgcctc | ccattctact | caaagtcata | cctctgttca | cagagatgga | 1200 |
| tgcattgttct | gattgcctct | ttggagaagc | tcatcagaaa | ctcaaaagaa | ggccactgtt | 1260 |
| tgtctcacct | acccatgacc | tgaagcccc | ccctgagtgg | tccccacctt | tctggacgga | 1320 |
| accacgtact | ttttacatac | attgattcat | gtctcacgtc | tccctaaaaa | tgcgtaagac | 1380 |
| caagctgtgc | cctgaccacc | ctgggcccc | gtcgtcagga | cctcctgagg | ctttggcaaa | 1440 |
| taaacctcct | aaaatgaaaa | aaaaaaagg | | | | 1469 |

<210> 39

<211> 5543

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7504930CB1

<400> 39

| | | | | | | |
|-------------|------------|------------|-------------|------------|-------------|------|
| ctggagcagc | ggcgccccgg | agccgagctt | gcagcgagg | accggctgag | gcgcgcggga | 60 |
| gggaaggagg | caagggctcc | gcggcgctgt | cgcgctgccg | ctcactctcg | gggaagagat | 120 |
| ggcgccggag | cggggagccc | ggcgactcct | cagcaccccc | tccttctggc | tctactgcct | 180 |
| gctgctgctc | gggcgcggg | cgccggggcg | cgcgccggcc | aggagcggct | ccgcgcggca | 240 |
| gtccccagga | gccagcattc | gaacgttcac | tccattttat | tttctggtgg | agccgggtgga | 300 |
| tacactctca | gttagaggct | cttctgttat | attaaactgt | tcagcatatt | ctgagccttc | 360 |
| tccaaaaatt | gaatggaaaa | aagatggaac | ttttttaaac | ttagtatcag | atgatcgacg | 420 |
| ccagcttctc | ccggatggat | ctttatttat | cagcaatgtg | gtgcattcca | aacacaataa | 480 |
| acctgatgaa | ggttattatc | agtgtgtggc | cactgttgag | agtcttggaa | ctattatcag | 540 |
| tagaacagcg | aagctcatag | tagcaggtct | tccaagattt | accagccaac | cagaaccttc | 600 |
| ctcagtttat | gctgggaaca | atgcaattct | gaattgtgaa | gttaatgcag | atttgggtccc | 660 |
| atttgtgagg | tgggaacaga | acagacaacc | ccttcttctg | gatgatagag | ttatcaaact | 720 |
| tccaagtgga | atgctggtta | tcagcaatgc | aactgaagga | gatggcgggc | tttatcgctg | 780 |
| cgtagtggaa | agtgggtggc | caccaaagta | tagtgatgaa | gttgaattga | aggttcttcc | 840 |
| agatcctgag | gtgatatcag | acttgggtat | tttgaacacg | ccttctccct | tagtcagagt | 900 |
| cattggctcag | gatgtagtgt | tgccatgtgt | tgccttcagga | cttcctactc | caaccattaa | 960 |
| atggatgaaa | aatgaggagg | cacttgacac | agaaagctct | gaaagattgg | tattgctggc | 1020 |
| aggtggtagc | ctggagatca | gtgatgttac | tgaggatgat | gctgggactt | atttttgtat | 1080 |
| agctgataat | ggaaatgaga | caattgaagc | tcaagcagag | cttacagtgc | aagctcaacc | 1140 |
| tgaattcctg | aagcagccta | ctaatatata | tgtcacgaa | tctatggata | ttgtatttga | 1200 |
| atgtgaagtg | actggaaaac | caactccaac | tgtgaagtgg | gtcaaaaatg | gggatattgg | 1260 |
| tatcccaagt | gattatttta | agattgtaaa | ggaacataat | cttcaagttt | tgggtctggg | 1320 |
| gaaatcagat | gaagggttct | atcagtgcat | tgcgtgaaa | gatgttggaa | atgcacaagc | 1380 |
| tggagcccaa | ctgataatcc | ttgaacatgc | accagccaca | acgggaccac | tgccttcagc | 1440 |
| tcctcgggat | gtcgtggcct | ccctgggtct | taccgcgttc | atcaaatgta | cgtggcggac | 1500 |
| acctgcatac | gacccctcac | gagacaacct | tacctactct | gtgttctaca | ccaaggaagg | 1560 |
| gattgctagg | gaacgtgttg | agaataccag | tcacccagga | gagatgcaag | taaccattca | 1620 |
| aaacctaatg | ccagcgaccg | tgtacatctt | tagagtatat | gctcaaaata | agcatggctc | 1680 |
| aggagagagt | tcagctccac | tgcgagtaga | aacacaacct | gaggttcagc | tccttgcccc | 1740 |
| agcacctaac | cttcgtgcat | atgcagcttc | gcctacctcc | atcactgtta | cgtgggaaac | 1800 |
| accagtgtct | ggcaatgggg | aaattcagaa | ttataaattg | tactacatgg | aaaaggggac | 1860 |
| tgataaagaa | caggatgttg | atgtttcaag | tcactcttac | accattaatg | ggttgaaaaa | 1920 |
| atatacagag | tatagtttcc | gagtggtggc | ctacaataaa | catggtcctg | gagtttccac | 1980 |
| accagatgtt | gctgttcgaa | cattgtcaga | tgttcccagt | gctgctcctc | agaatctgtc | 2040 |
| cttgggaagt | agaaattcaa | agagtattat | gattcactgg | cagccacctg | ctccagccac | 2100 |
| acaaaatggg | cagattactg | gctacaagat | tcgctaccga | aaggcctccc | gaaagagtga | 2160 |
| tgtcactgag | accttggtaa | gcgggacaca | gctgtctcag | ctgattgaag | gtccttgatc | 2220 |
| ggggactgag | tataatttcc | gagtggtcgt | tctacaacat | aatgggtacg | gcccggcaac | 2280 |
| tgactggctg | tctgctgaaa | cttttgaaag | tgacctagat | gaaactcgtg | ttcctgaagt | 2340 |
| gcctagctct | cttcacgtac | gcccgcctcg | tactagcatc | gtagtgcgtc | ggactcctcc | 2400 |
| agagaatcag | aacattgtgg | tcagaggtta | cgccattggg | tatggcattg | gcagccctca | 2460 |
| tgcccagacc | atcaaagtgg | actataaaca | gcgctattac | accattgaaa | atctggatcc | 2520 |
| cagctctcac | tatgtgatta | ccctgaaagc | atttaataac | gtgggtgaag | gcatccccct | 2580 |
| gtatgagagt | gctgtgacac | ggcctcacac | agtgccagat | cccactccca | tgatgccacc | 2640 |

```

agtgggagtt caggcttcca ttctgagtca tgacaccatc aggattacgt gggcagacaa 2700
ctcgctgccc aagcaccaga agattacaga ctcccgatac tacaccgtcc gatggaaaac 2760
caacatccca gcaaacacca agtacaagaa tgcaaagtca accactttga gttattttggt 2820
gactggttta aagccgaata cactctatga attctctgtg atggtgacca aagggtcgaag 2880
atcaagtaca tggagtatga cagcccatgg gaccaccttt gaattagttc cgacttctcc 2940
acccaaggat gtgactgttg tgagtaaaga ggggaaacct aagaccataa ttgtgaattg 3000
gcagcctccc tccgaagcca atggcaaaat tacaggttac atcatatatt acagtacaga 3060
tgtgaatgca gagatacatg actgggttat tgagcctgtt gtgggaaaca gactgactca 3120
ccagatacaa gagttaactc ttgacacacc atactacttc aaaatccagg caccggaactc 3180
aaagggcatg ggacccatgt ctgaagctgt ccaattcaga acacctaag cggactcctc 3240
tgataaaatg cctaattgat aagcctcagg gtctggaggg aaaggaagcc ggctgccaga 3300
cctaggatcc gactacaaac ctccaatgag cggcagtaac agccctcatg ggagccccac 3360
ctctcctctg gacagtaata tgctgctggt cataattgtt tctgtaggcg tcatcaccat 3420
cgtgggtggt gtgattatcg ctgtcttttg taccctgtcg accacctctc accgaaaaa 3480
gaaacagagct gctgcaaat cagtgaattg ctctcataag tacaaagga attccaaaga 3540
tgtgaaacct ccagatctct ggatccatca tgagagactg gagctgaaac ccattgataa 3600
gtctccagac ccaaacccca tcatgactga tactccaatt cctcgcaact ctcaagatat 3660
cacaccagtt gacaactcca tggacagcaa tatccatcaa aggcgaaatt catacagagg 3720
gcatgagtca gaggacagca tgtctacact ggctggaagg cgaggaatga gacaaaaaat 3780
gatgatgccc ttgtactccc agccaccccc gcctgtgatt agtgcccatc ccatccattc 3840
cctcgataac cctcaccatc atttccactc cagcagcctc gcttctccag ctgcgagta 3900
tctctaccac ccgggcagcc catggcccat tggcacatcc atgtcccttt cagacagggc 3960
caattccaca gaatccgttc gaaatacccc cagcactgac accatgccag cctcttcgtc 4020
tcaaacatgc tgcactgatc accaggaccc tgaagggtgt accagctcct cttacttggc 4080
cagctcccaa gaggaagatt caggccagag tcttcccact gcccatgttc gcccttccca 4140
cccattgaag agcttcgccg tgccagcaat ccgcctcca ggacctcca cctatgatcc 4200
tgcatggcca agcacaccat tactgtccca cgaagctctg aaccatcaca ttcactcagt 4260
gaagacagcc tccatcgga ctctaggaag gagccggcct cctatgccag tgggtgttcc 4320
cagtgcctct gaagtgcagg agaccacaag gatgttgga gactccgaga gtagctatga 4380
accagatgag ctgaccaaag agatggccca cctggaagga ctaatgaagg acctaaacgc 4440
tatcacaaca gcatgacgac ctccaccagg acctgacttc aaacctgagt ctggaagtct 4500
tggaacttaa cccttgaaaa caaggaattg tacagagtac gagaggacag cacttgagaa 4560
cacagaatga gccagcagac tggccagcgc ctctgtgtag ggctggctcc aggcattggc 4620
acctgccttc ccctggtcag cctggaagaa gcctgtgtcg aggcagcttc cctttgcctg 4680
ctgatattct gcaggactgg gcaccatggg ccaaaatttt gtgtccaggg aagagggcag 4740
aagtgcaccc tgcatttcac tttgtggtca ggccgtgtct ttgtgctgtg actgcatcac 4800
ctttatggag tgtagacatt ggcatttatg tacaatttta tttgtgtctt attttatttt 4860
accttcaaaa acaaaaacgc catccaaaac caaggaagtc cttggtgttc tccacaagtg 4920
gttgacattt gactgcttgt tccaattatg tatggaaagt ctttgacagt gtgggtcgtt 4980
cctgggggtt gcttgttttt tggtttcatt tttatttttt aattctgagt cattgcatcc 5040
tctaccagct gttaatccat cactctgagg gggaggaaat gttgcattgc tgtttgtaag 5100
ctttttttat tattttttta ttataattat taaaggcctg actctttcct ctcatcactg 5160
tgagattaca gatctatttg aattgaatga aatgtaacat tgaaaagact tgtttgttgc 5220
tttctgtgca gtttcagtat tggggcgggt ggggggctgg gacttggtta taggaaatgg 5280
aggaactgct gaggtcctgt gaatgtttct gtcattgtac tttcttcag aagcctgcag 5340
agaatggaag catcttcttt attgtccttt cctggcatgt ccaccttat tgtcactacg 5400
ttgcaacgtg gagtttgatt tggatctggt tttaaaattc ttctatgcaa tacgtgggtt 5460
tgaggattta gcgccctga tgtcttggtc atagcctggt aagaatgtcc atgctgagga 5520
gccacatggt gtattttctaa ctg 5543

```

<210> 40

<211> 4293

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6610456CB1

<400> 40

```

agcggggcga agagtagcgg taggtcggcg ggacttccgt gttggcggga ttctgaacgc 60
tgccatggct cagaccgtgc agaattgtac attgtcgctc actctgccca tcacgtgcc 120
catttgcttg ggggaaggtag gtcagcctgt catatgcac aacaaccatg tattttgttc 180

```

| | | | | | | |
|-------------|-------------|------------|-------------|-------------|-------------|------|
| gatttgtatt | gatttgtggt | tgaagaataa | tagccagtgt | ccagcttgca | gagtccecat | 240 |
| cactcctgaa | aatccttgca | aagaaattat | aggaggaaca | agtgaagtg | aacctatgct | 300 |
| aagccatagc | gtcaggaagc | atcttcggaa | aactagactt | gaattactac | acaaagaata | 360 |
| tgaggacgaa | atagattgtt | tacagaaaga | agtagaagag | cttaagagta | aaaatctcag | 420 |
| cttggagtca | cagatcaaaa | ctattctgga | tcctttaacc | ttggtgcagg | gcaaccaaaa | 480 |
| tgaagacaaa | catctagtca | cagataatcc | aagtaaaatt | aaccagaaa | ctgtagcaga | 540 |
| gtggaagaaa | aaactcagaa | cagctaata | aatctatgaa | aaagtgaag | atgatgtgga | 600 |
| taagctaaag | gaggcaata | aaaaattgaa | attggaaaat | ggtggtctgg | tgaggagaa | 660 |
| tttacgactg | aaggctgaag | ttgataacag | atcacctcaa | aagtttgga | ggtttgagc | 720 |
| tgctgctctt | cagtcocaa | tgaacagta | tgagcgtgaa | accaatcgcc | tcaagaaagc | 780 |
| cctggaacga | agtgataagt | atatagagga | actagaatct | caagttgcac | agctaaaaaa | 840 |
| ttcaagtga | gagaaagaag | ctatgaattc | catttgccag | acagcacttt | ctgcagatgg | 900 |
| caaagggagc | aaaggcagtg | aggaggatgt | ggtgtcaaag | aatcaaggcg | atagtgccag | 960 |
| aaagcagcct | ggctcatcca | cctccagttc | ttctcaccta | gcgaagcctt | ccagcagcag | 1020 |
| actgtgtgac | accagttctg | caaggcagga | aagtacagc | aaagcagacc | ttactgttcc | 1080 |
| taagaacaaa | gacctatata | aagaacaggt | agaagtaatg | ttagatgtga | cagatacaag | 1140 |
| tatggatact | tatttggaaa | gagaatgggg | gaataaacca | agtgactgtg | taccctacaa | 1200 |
| agatgaagaa | ctttatgata | ttccagctcc | ttgtactcct | ttgtccctta | gttgccctca | 1260 |
| gctcagtact | ccagaaaata | gagagagctc | tgtggtccaa | gcaggagggtt | ccaaaaagca | 1320 |
| ctcaaaccat | ctcagaaaat | tggtgtttga | tgatttttgt | gattcttcaa | atgtttctaa | 1380 |
| taaagattct | tcagaagatg | atataagtag | aagtgaaaat | gaaaagaaat | cagaattgtt | 1440 |
| ttcttcccca | aagacaggat | tttgggactg | ttgtccaca | agctatgcc | aaaacttaga | 1500 |
| ttttgaaagt | tcagagggga | acacgatagc | aaattctgtt | ggagaaatat | cttcaaaatt | 1560 |
| gagtggagaa | tcaggcttat | gtttatccaa | aagggttgaat | tctattcgct | cttttgaaat | 1620 |
| gaaccggaca | agaacatcca | gtgaagcatc | gatggatgct | gcttaccttg | acaaaatctc | 1680 |
| tgagttggat | tcaatgatgt | cagagtcaga | caacagcaag | agcccttgta | ataacggttt | 1740 |
| taagtctact | gatttggatg | ggttatcaaa | gtcatctcaa | ggcagtgaat | ttcttgagga | 1800 |
| acctgataag | ttggaagaaa | aaactgagct | aaacctttcc | aaagggtctc | taactaatga | 1860 |
| tcagttagaa | aatggaagtg | aatggaacc | cacttctttt | tttctcctct | ctccatctga | 1920 |
| ccaagaaatg | aatgaagatt | tttactcca | ttccagttct | tgtccagtaa | ctaataagac | 1980 |
| caaaccacca | agctgcttgt | ttcagacaga | gttttccag | ggcattttgt | taagcagttc | 2040 |
| acatcgacta | tttgaagatc | aaagatttgg | gtcatctttg | tttaagatgt | cctcagagat | 2100 |
| gcacagtctt | cataaccacc | ttcagtcctc | ttggtctact | tcctttgtgc | tgaaaagag | 2160 |
| gaataaaaaa | gtgaatcaat | caacaaaag | aaaaatccag | agcagccttt | ccagtgccag | 2220 |
| cccataaaaa | gcaactaaaa | gttgactcat | tagaaagggtg | tcatttgtgg | ttttgtcctg | 2280 |
| agagaaatag | aaaagttggt | aaagttacct | tttttctca | taaaagttct | atacaaattg | 2340 |
| gaattgataa | tcttttagtca | agtatcaagt | caggatgggtg | gattaacctg | taccagaat | 2400 |
| acttattgtt | catttttga | agactttgtt | cttttcat | ttatttggga | gtctttgtga | 2460 |
| ccagagaagt | tagggaggag | gttatttttg | tgttttgggg | ttggttggtt | ggttggtttt | 2520 |
| gtttttgggt | ttgttttttt | actgaatttg | atatgtatct | cggttggata | tacattgttt | 2580 |
| tttttaaaaa | tgttatttta | ctgttagata | cagtggcctg | ttgataagcc | ccacttgtct | 2640 |
| tcagaacttg | gatttcttaa | ataaaacttt | tagtgttgct | tatacactgc | tcaataagac | 2700 |
| acttgagttt | aagcttttcc | cagggtggaa | attattttac | ctgtcccttt | ttatttatgt | 2760 |
| ttagtgatgg | cctagttttt | ctgcagggcc | atgatggaga | aatagcactc | tagccttagt | 2820 |
| ccaatattga | tttactttct | ttttttagg | tttatgtata | tgtttgcat | tttttagcatt | 2880 |
| gtgttttgtc | cagttttgtg | aaaatgttct | gctagtatga | aagaaaacat | tttctatatg | 2940 |
| aagacatttg | ttttatgtta | ggtagcttac | attttctcct | ctgcgtgtgt | gtgtatgtgt | 3000 |
| gtaaaatcag | aaatttagca | tactatggaa | agaaggcatg | gagcacttgg | gttttagagga | 3060 |
| acctaaaaca | tcatagcttc | attgttccag | atgtaacagg | tttgaaagag | ctcatcgcca | 3120 |
| agttcttgat | ccacttgcac | tccaggggag | ttttcttttg | agtagtatgt | ttcttgtttg | 3180 |
| catgttctctg | ttctttgttg | aaactatgca | tggtagcatt | tttgcttgct | gtgttttcca | 3240 |
| tacttaagaa | aaagaggttt | cagttggctg | atagaatatc | ttttatgtag | gacaaaactt | 3300 |
| ttctgtgaag | agtgttgagg | gggtgaagat | aggtaagagg | taagcacaat | ttttaattta | 3360 |
| ggctctgaaa | aagtgtattg | ttctaaacgt | atttgggtatg | cctatatagg | tcttttaaaaa | 3420 |
| tgggtatgta | tgtgttttaa | tgtgcactga | acattttaca | ttaatattgt | actgttttac | 3480 |
| attaatactg | catgcttttc | tatgtgaatt | gaataaagaa | tgtcataagc | actgtgaaaa | 3540 |
| aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaatagaaaa | 3600 |
| gaaaaaaaaa | aaataaaaaa | ataaaaaaaa | aaaaaa | aaataatggg | ttggaaaaat | 3660 |
| aagagaagga | cttgattagg | gaccagaatg | tcggggaaaa | gggaaaaagc | gaagggtaag | 3720 |
| agcgcaggat | aagttccaaa | gaattcgggg | aggaaaagag | aatatgggaa | ggggtgaagc | 3780 |
| aaagaaagaa | gacgaaggga | tcagaaagca | gcaggtgaag | tgaggaaatg | aagagcagaa | 3840 |
| ggtcgataga | acgtgagcaa | agaatagata | gaatgtaatt | aaagggatga | ggaaagagag | 3900 |
| atggggagga | tgaagagaat | tgagattgtg | gtgataacga | gagaaagtga | ggagatggtc | 3960 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| agagagagcg | tgaaggtaga | gagttgtgga | cgtgaggagg | aagtggaggc | gcaattgtgc | 4020 |
| gaagtgaagt | agagaagtgc | gacgagtcgg | aaaagagatg | gagagagaga | gcgagagagg | 4080 |
| agaaagaaat | agaaagggtt | gagaaggaga | gacttaggta | gaagagagga | ggagcgaaga | 4140 |
| aggaggagag | gagagatggc | gagcgtaggg | ggaagaaaga | gcaagcagtg | aacgtgggga | 4200 |
| taatggagag | agagaaagtg | aagagaggta | gaggggagga | gtggatcaga | taatgaagag | 4260 |
| aggggggaac | gaggagagag | aggaagaggt | ggg | | | 4293 |

<210> 41

<211> 4777

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503573CB1

<400> 41

| | | | | | | |
|------------|-------------|------------|-------------|------------|-------------|------|
| atgggagacg | taaaagcggt | gctgtttgtc | gctgctgccc | gggccagacg | tctaggagga | 60 |
| gccgctgcat | ccgagtccct | ggctgtctcc | gaagccttct | gcagggtccg | aagctgccag | 120 |
| cccaaaaagt | gtgcaggccc | ccagcgggtg | ctgaacccag | tgcttgacgt | gcccagtcct | 180 |
| agccccagcg | tgaggaaagc | acaggtgtcc | ctcaactggc | agccactgac | gctccaggag | 240 |
| gccagagctc | tactgaagcg | gcggcggccc | cggggggccag | ggggccgggg | actactgaga | 300 |
| aggaggcccc | cacagcgtgc | ccccgctggc | aaggcccccg | tcctgtgtcc | cttgatctgt | 360 |
| cacaatggcg | gtgtgtgctg | gaagcctgac | cgctgcctct | gtcccccgga | cttcgctggc | 420 |
| aagttctgcc | agttgcactc | ctcgggcgcc | cggcccccg | ccccggctgt | accaggcctc | 480 |
| accgcctccg | tgtacactat | gccactggcc | aaccaccgcg | acgacgagca | cggcgtggca | 540 |
| tctatgggtg | gcgtccacgt | ggagcaccgc | caggaggcgt | cgggtgggtg | gcaccagggt | 600 |
| gagcgtgtgt | ctggcccttg | ggaggaggcg | gacgctgagg | cggtgggcgc | ggcggaaagc | 660 |
| gcggcgcggg | cggaggcggc | agcgccctac | acgggtgttg | cacagagcgc | gcccgcgggag | 720 |
| gacggctact | cagatgcctc | gggcttcggg | tactgctttc | gggagctgcg | cggaggcgaa | 780 |
| tgcgcgctcc | cgctgcccgg | gctccggacg | caggaggctc | gctgcccagg | ggccggcttg | 840 |
| gcctggggcg | ttcacgactg | tcagctgtgc | tcagagcgcc | tggggaaact | cgaaagagtg | 900 |
| agcgccccag | atggaccttg | tccaaccggc | tttgaaagag | ttaatgggtc | ctgcgaagat | 960 |
| gtggatgagt | gcgcgactgg | cgggcgctgc | cagcacggcg | agtgtgcaaa | cacgcgcggc | 1020 |
| gggtacacgt | gtgtgtgccc | cgacggcttt | ctgctcgact | cgccccgcag | cagctgcata | 1080 |
| tccaacacag | tgatctcaga | ggccaaaggg | ccctgcttcc | gcgtgctccg | cgacggcggc | 1140 |
| tgttcgctgc | ccattctgcg | gaacatcact | aaacagatct | gctgctgcag | ccgcgtaggg | 1200 |
| aaggcctggg | gccgggggctg | ccagctctgc | ccacccttcg | gctcagaggg | tttcggggag | 1260 |
| atctgcccgg | ctggtcctgg | ttaccactac | tcggcctccg | acctccgcta | caacaccaga | 1320 |
| ccctcgcccg | aggagccacc | ccgagtgtca | ctcagccagc | ctcgtaacct | gccaccgacc | 1380 |
| tctcgcccat | ctgcaggctt | tctgcccacc | catcgcttgg | agccccggcc | tgaaccccg | 1440 |
| cccgatcccc | ggcccggccc | tgagcttccc | ttgcccagca | tccttgcttg | gactgggtcc | 1500 |
| gagattcctg | aatcagggtcc | ctcctccggc | atgtgtcagc | gcaaccccca | ggtctgcggc | 1560 |
| ccaggacgct | gcatttcccc | gcccagcggc | tacacctgcg | cttgcgactc | tggtctccgg | 1620 |
| ctcagcccc | agggcaccgc | atgcattgat | gtggacgaat | gtcgccgcgt | gccccgcgcc | 1680 |
| tgtgtctccg | ggcgctgcga | gaactcacca | ggcagcttcc | gctgcgtgtg | cggcccgggc | 1740 |
| ttccgagccg | gcccacgggc | tgcggaatgc | ctggatgtgg | acgagtgcga | ccgcgtgcgc | 1800 |
| ccgcctgtgt | acctcgggcg | ctgcgagaac | acgccaggca | gcttctctgt | cgtgtgcccc | 1860 |
| gccgggtacc | aggctgcacc | gcacggagcc | agctgccagg | atgtggatga | atgcaccag | 1920 |
| agcccaggcc | tgtgtggccg | aggggcctgc | agaacacctg | ctggctcttt | ccgctgtgtt | 1980 |
| tgcccggctg | gcttccgggg | ctcggcgtgt | gaagaggatg | tggatgagtg | tgcccaggag | 2040 |
| ccgcgcctcc | gtgggcccgg | ccgctgtgac | aacacggcag | gctcctttca | ctgtgcttcc | 2100 |
| cctgctgggt | tccgctcccg | agggcccggg | gccccctgcc | aagatgtgga | tgagtgtgcc | 2160 |
| cgaagccccc | caccctgcac | ctacggccgg | tgtgagaaca | cagaaggcag | cttcagtggt | 2220 |
| gtctgcccc | tgggcttcca | acccaacact | gctggctccg | agtgcgagga | tgtggatgag | 2280 |
| tgtgagaacc | acctcgcatg | ccctgggcag | gagtgtgtga | actcgcccgg | ctccttccag | 2340 |
| tgcaggacct | gtccttcttg | ccaccacctg | caccgtggca | gatgcactga | tgtggacgaa | 2400 |
| tgagtttcgg | gtgcccctcc | ctgtggtccc | cacggccact | gcactaacac | cgaaggctcc | 2460 |
| ttccgctgca | gctgcgcgcc | aggctaccgg | gcgcgcctgg | gtcggcccgg | gccctgcgca | 2520 |
| gacgtgaacg | agtgcctgga | gggcgatttc | tgcttccctc | acggcgagtg | cctcaacact | 2580 |
| gacggctcct | ttgcctgtac | ttgtgcccct | ggctaccgac | ccggaccccc | cggagcctct | 2640 |
| tgcttcgacg | ttgacgagtg | cagcgaggag | gacctttgcc | agagcggcat | ctgtaccaac | 2700 |
| accgacggct | ccttcgagtg | catctgtcct | cggggacacc | gcgctggccc | ggacctcgcc | 2760 |

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tcctgcctcg | acgtggacga | atgtcgcgag | cgaggcccag | ccctgtgccc | gtcgcagcgc | 2820 |
| tgtgagaact | ctccccggctc | ctaccgcgtgt | gtcccggaact | gcgatccctgg | gtaccacgcg | 2880 |
| ggccccgagg | gcacctgtga | cgatgtggat | gagtgcctaag | aatatgggtcc | cgagatttgt | 2940 |
| ggagcccagc | gttgtgagaa | caccctggc | tcctaccgct | gcacaccagc | ctgtgaccct | 3000 |
| ggctatcagc | ccacgccagg | ggcgcatgc | caggatgtgg | acgaatgccg | gaaccggtcc | 3060 |
| ttctgcggtg | cccacgccgt | gtgccagAAC | ctgcccggct | ccttccagtg | cctctgtgac | 3120 |
| cagggttacg | agggggcacg | ggatgggctg | cactgcgtgg | atgtgaacga | gtgtgaaaca | 3180 |
| ctacaggggtg | tatgtggagc | tgcctgtgt | gaaaatgtcg | aaggctcctt | cctctgtgtc | 3240 |
| tgccccaaac | gcccgaaga | gtttgacccc | atgactggac | gctgtgttcc | cccacgaact | 3300 |
| tctgctgacg | tggacgaatg | tcagctcttc | cgagaccagg | tgtgcaagag | tggcgtgtgt | 3360 |
| gtgaacacgg | ccccgggcta | ctcatgctat | tgcagcaacg | gctactacta | ccacacacag | 3420 |
| cggctggagt | gcacgcacaa | tgacgagtgc | gccgatgagg | aaccggcctg | tgagggcggc | 3480 |
| cgctgtgtca | acactgtggg | ctcttatcac | tgtacctgcg | agccccact | ggtgctggat | 3540 |
| ggctcgcagc | gccgctgcgt | ctccaacgag | agccagagcc | tcgatgacaa | tctgggagtg | 3600 |
| tgttggcagg | aagtgggggc | tgacctcgtg | tgcagccacc | ctcggctgga | ccgtcaggcc | 3660 |
| acctacacag | agtgtgctg | cctgtatgga | gaggcctggg | gcattggactg | cgccctctgc | 3720 |
| cctgcgcagg | actcagatga | cttcgaggcc | ctgtgcaatg | tgtacgccc | ccccgcatat | 3780 |
| agccccccgc | gaccaggtgg | ctttggactc | ccctacgagt | acggcccaga | cttaggtcca | 3840 |
| ccttaccagg | gcctcccata | tgggcctgag | ttgtaccac | cacctgcgct | accctacgac | 3900 |
| ccctaccac | cgccacctgg | gcccttcgcc | cgccgggagg | ctccttatgg | ggcaccgcc | 3960 |
| ttcgacatgc | cagactttga | ggacgatggt | ggccctatg | gcgaatctga | ggctcctgcg | 4020 |
| ccacctgggc | cgggcaccgc | ctggccctat | cggtcgccgg | acaccgcgcg | ctccttccca | 4080 |
| gagccccagg | agcctcctga | aggtggaagc | tatgctgggt | ccctggctga | gccctacgag | 4140 |
| gagctggagg | cggaggagtg | cgggatcctg | gacggctgca | ccaacggccg | ctgcgtgcgc | 4200 |
| gtccccgaag | gcttcacctg | ccgttgcttc | gacggctacc | gcctggacat | gaccgcgatg | 4260 |
| gcctgcgttg | acatcaacga | gtgtgatgag | gccgaggctg | cctccccgct | gtgcgtcaac | 4320 |
| gcgcgttgcc | tcaacacgga | tggctccttc | cgctgcattc | gccgcccggg | attcgcacc | 4380 |
| acgcacacgc | cgcaccactg | tgcgcgccga | cggcccccgg | cctgagccct | ggcaccgcgt | 4440 |
| ggccgcccac | cgcgcgccgc | cactcggggc | ccctgcgcgc | catactgcag | cccgttatg | 4500 |
| cgtatgtgca | cggggccgcc | cgcctggacc | tggagaaggg | acctacggac | gcctggaagc | 4560 |
| tgcgacgccc | tgcactgctc | ccgcctccac | cagcgcctcc | cactgatgtc | gtgggtcccgc | 4620 |
| gcctggccca | ggggccctt | tacatgccct | ctccctttta | taaaattttc | cattaaaaac | 4680 |
| cacctatttt | ctaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | 4740 |
| aaaaaaaaaa | aaaaaaaaaa | aaaaaaaaaa | aagaggg | | | 4777 |

<210> 42

<211> 1463

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7505057CB1

<400> 42

| | | | | | | |
|------------|-------------|-------------|-------------|-------------|-------------|------|
| gtctgcagca | gcattttaaat | tctgggaggg | cttgggttgct | agcagcagca | ggaggaggga | 60 |
| gagcacagca | tcgtcgggac | cagactcgctc | tcaggccagt | tgcagccttc | tcagccaaac | 120 |
| gccgaccaag | gaaaactcac | taccatgaga | attgcagtga | tttgcttttg | cctcctagga | 180 |
| atcacctgtg | ccataccagt | taaacaggct | gattctggaa | gttctgagga | aaagcagacc | 240 |
| cttccaagta | agtccaacga | aagccatgac | cacatggatg | atatggatga | tgaagatgat | 300 |
| gacgaccatg | tggacagcca | ggactccatt | gactcgaacg | actctgatga | tgtagatgac | 360 |
| actgatgatt | ctcaccagtc | tgatgagtct | caccattctg | atgaatctga | tgaactggct | 420 |
| actgattttc | ccacggacct | gccagcaacc | gaagttttca | ctccagttgt | ccccacagta | 480 |
| gacacatatg | atggccgagg | tgatagtgtg | gtttatggac | tgagggtcaaa | atctaagaag | 540 |
| tttcgcagac | ctgacatcca | gtaccctgat | gctacagacg | aggacatcac | ctcacacatg | 600 |
| gaaagcgagg | agttgaatgg | tgcatacaag | gccatccccg | ttgcccagga | cctgaacgcg | 660 |
| ccttctgatt | gggacagccg | tgggaaggac | agttatgaaa | cgagtcagct | ggatgaccag | 720 |
| agtgttgaaa | cccacagcca | caagcagtc | agattatata | agcggaaagc | taattgatgag | 780 |
| agcaatgagc | attccgatg | gattgatagt | caggaaacttt | ccaaagtcag | ccgtgaattc | 840 |
| cacagccatg | aatttcacag | ccatgaagat | atgctgggtt | tagaccccaa | aagtaaggaa | 900 |
| gaagataaac | acctgaaatt | tcgtattttc | catgaattag | atagtgcac | ttctgagggt | 960 |
| aattaaaagg | agaaaaaata | caattttctc | ctttgcattt | agtcaaaaga | aaaaatgctt | 1020 |
| tatagcaaaa | tgaaagagaa | catgaaatgc | ttctttctca | gtttattgggt | tgaatgtgta | 1080 |

| | | | | | | |
|-------------|------------|-------------|------------|------------|------------|------|
| tctatattgag | tctggaaata | actgatgtgt | ttgataatta | gtttagtttg | tgggcttcat | 1140 |
| ggaaactccc | tgtaactaa | aagcttcagg | gttatgtcta | tggacaaact | atagaaggaa | 1200 |
| tgcaaacctat | cccagtat | taataattgg | taaacacacc | aggataggaa | attaaggagg | 1260 |
| cgcaacgcaa | aactaaaccc | cccagaaaag | aggatatacc | atgtaggccc | aaaaaccgtg | 1320 |
| ggtgagtatg | aaggggataa | ttggaggggga | gaccaaaggg | ggtgggtaag | ccttaaccag | 1380 |
| ggaggaaagc | ggaaggaaga | aaggggcgcc | ccaggcaacc | cccagggggg | ggaaaaagca | 1440 |
| ggccgcccc | gagcaccgcg | gcc | | | | 1463 |

<210> 43

<211> 1259

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 90116002CB1

<400> 43

| | | | | | | |
|-------------|-------------|-------------|------------|------------|-------------|------|
| caagaccctg | actcaaagaa | cacctctcac | tacattcaga | gtctgtcatc | tgaaccatga | 60 |
| ggatctgggtg | gcttctgctt | gccattgaaa | tctgcacagg | gaacataaac | tcacaggaca | 120 |
| cctgcaggca | agggcacccct | ggaatccctg | ggaaccccg | tcacaatggg | ctgcctggaa | 180 |
| gagatggacg | agacggagcg | aagggtgaca | aaggcgatgc | aggagaacca | ggacgtcctg | 240 |
| gcagcccggg | gaaggatggg | acgagtggag | agaagggaga | acgaggagca | gatggaaaag | 300 |
| ttgaagcaaa | aggcatcaaa | ggtgatcaag | gctcaagagg | atccccagga | aaacatggcc | 360 |
| ccaaggggct | tgcagggccc | atgggagaga | aaggcctccg | aggagagact | gggcctcagg | 420 |
| ggcagaaggg | gaataagggt | gacgtgggtc | ccactgggtc | tgaggggcca | aggggcaaca | 480 |
| ttgggccttt | gggcccact | ggtttaccgg | gccccatggg | ccctattgga | aagcctgggtc | 540 |
| ccaagggaga | agctggaccc | acggggcccc | agggtgagcc | aggagtccgg | ggaataagag | 600 |
| gctggaaagg | agatcgagga | gagaaaaggga | aaatcgggtg | gactctagtc | ttgccaaaaa | 660 |
| gtgctttcac | tgtggggctc | acggtgctga | gcaagtttcc | ttcttcagat | gtgcccatta | 720 |
| aatttgataa | gatcctgtat | aacgaattca | accattatga | tacagcagcg | gggaaattca | 780 |
| cgtgccacat | tgtctggggtc | tattacttca | cctaccacat | cactgttttc | tccaggaatg | 840 |
| ttcaggtgtc | tttgggtcaa | aatggagtaa | aaatactgca | caccaagat | gcttacatga | 900 |
| gctctgagga | ccaggcctct | ggcggcctct | tcctgcagct | gaagctcggg | gatgaggtgt | 960 |
| ggctgcagggt | gacaggagga | gagagggttca | atggcttgtt | tgctgatgag | gacgatgaca | 1020 |
| caactttcac | agggttccct | ctgttcagca | gcccgtgaca | gaggagagtt | taaaaatccg | 1080 |
| ccacaccatc | catcagaatc | agcttgggat | gaacttattc | agatgggttt | actttattaa | 1140 |
| ttcctccaat | tattacaata | atcataaaaa | ggtgaaaatg | gaaaagttat | tcccaaaact | 1200 |
| gattctgtgt | aacttactat | ttttccagga | gtaaatat | aaaatagcaa | aaaaaaaa | 1259 |

<210> 44

<211> 3548

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 039283CB1

<400> 44

| | | | | | | |
|------------|------------|-------------|-------------|------------|------------|-----|
| taactagtaa | atggaaatga | gtgtgtgggtg | gccttggggc | tcctcagtct | gagctcagcc | 60 |
| catctggggc | tcctcctagg | ttctgggctt | ccatctgctt | ccccctcctt | ttgggtccct | 120 |
| ctgagaagcc | tcccaagatc | gagctgtgta | cagagcgagt | cctcgctggc | cactcggaag | 180 |
| gcggcggacg | tcgcgcgcgg | cgcactagcc | cagagcggcc | tgggtctccg | ccagtgggcg | 240 |
| gggcctctta | ccgggggccc | ggccgaggcg | tgccggcgagc | gctggagggg | tctcgtgggc | 300 |
| gctcccgtgg | ttgtggcccc | ggcggccgcg | agcgtgtgtg | tgtggcacga | tgctcgtggt | 360 |
| actcagaggc | gcgcgaggcg | gggcggccaa | cggggtcgcg | gggcgagcac | gcgcgccact | 420 |
| ggggcgccgg | acgttgcaag | ggggtcacat | ccggtagggg | gcgggcctgg | cggaccccg | 480 |
| aagtgtgggc | gcggctcgcg | tttaaccgcg | ggtggcgggc | gcgacggcgg | ccctggcagg | 540 |
| gaaggatgga | gacgctgaag | gataagaccc | tgcaggagct | ggaggagtgt | cagaatgact | 600 |
| cggaggcgat | tgaccagctg | gccctggagt | cccctgaggt | ccaggaccta | cagctggaac | 660 |
| gggagatggc | actggccacc | aaccggagcc | tggcagagcg | gaacttgag | ttccagggtc | 720 |
| ccctggagat | cagccgctca | aacctctcgg | atagatacca | ggagctccgg | aagctcgtgg | 780 |

```

agcgggtgcc  ggagcagaag  gcaaagctgg  agaaattttc  ttcagcactg  cagccaggga  840
ccttgttaga  ccttctgcag  gtggaaggca  tgaagatcga  agaagagtcc  gaggccatgg  900
ctgagaagtt  cctggagggc  gaggtgcccc  tggaaacggt  cctggagaat  ttttcctcca  960
tgaggatgct  gtccacacct  cgccgggttc  gcgtggaaaa  gctccaggaa  gtgggtgagga  1020
agcccagggc  ttcccaggag  ctggccggcg  atgccccctc  accccgtcca  ccaccccccg  1080
tgcgcccagt  cccccaggga  acacccccct  tggttgaaga  gcagccgcag  ccaccatcag  1140
ccatgcctcc  ctaccccttg  ccctacagcc  catccccag  cctgcctgtg  ggccccactg  1200
cccatggagc  cctgccaccg  gcccccttcc  cagtagtgct  ccagccctcc  ttctacagcg  1260
ggcctttggg  ccccatctac  ccggcagccc  agcttggacc  caggggtgct  gcgggttact  1320
cctggtcccc  acagaggagc  atgccacccc  ggccgggcta  tcctgggacc  ccaatgggtg  1380
cctctggggc  tgggtacccc  ttgcggggag  gcagggcccc  cagtcctggt  tatcctcaac  1440
agtccccata  ccccgcaca  ggaggaaaac  ctccctaccc  aatacagcct  cagctcccca  1500
gctttccagg  ccagccccag  ccctcagtgc  cctacagccc  atccccagc  ctgcctgtgg  1560
gccccactgc  ccatggagcc  ctgccaccgg  cccctttccc  agtagtgctc  cagccctcct  1620
tctacagcgg  gccctctggc  ccacttacc  cggcagcccc  gcttgagccc  aggggtgctg  1680
cgggttactc  ctggtcccca  cagaggagca  tgccaccccc  gccgggctat  cctgggaccc  1740
caatgggtgc  ctctgggcct  gggtagccct  tgccggggag  cagggccccc  agtcctggtt  1800
atcctcaaca  gtccccatac  cccgcaacag  gaggaaaacc  tccttacc  atacagcctc  1860
agctccccag  ctttccaggc  cagccccagc  cctcagtgc  cctgcagccc  ccttatcccc  1920
ccggggccgc  ccctccctat  gggttccccc  caccgcccgg  gcctgcctgg  cctgggtatt  1980
agacacactc  ctggccctcg  gccctcctc  agtcccaccc  acgctcaacc  ttgggccagg  2040
ccatcgccga  gattcgagg  tcaactggaa  gtggcggtgg  tgctccctgt  ggacttgcgt  2100
cggcacccag  aggccttgg  gggacaaggc  actgggcgg  gtgactggct  atggcacttg  2160
ctggccttct  ggccagaggt  cctcctggaa  gccctcctt  tcagtgcctg  gctgcagcgg  2220
tgctagccag  gctggagtcc  agcacctctc  gccggtgtct  gtgtgagtga  gcgcagtgtc  2280
gcatttatga  gcattctact  gcggctgcac  tcagggccat  ttgaaccag  tgcagagtta  2340
tctctatgag  gagggcccg  gtctccagcc  ccctccatct  cccctgggg  ttggcatatt  2400
tgacatggag  ccccatggca  agggatgcag  ctggtgggtg  ctgtccagtc  cattcaaggc  2460
tagcctgccc  atgcactgtg  gcaggggttg  gattggagg  agagaacaga  gttgggaaaa  2520
acaacagggt  tgagtcctat  aaagccataa  ttttaactcc  gtagctgatg  tcagacaagc  2580
ttgtcctatg  tcctatttga  gtggcagcag  cgccagcccc  gcaagaaggc  tgggggttgt  2640
caaggttgtc  cccagacctt  gcttgcagt  gttggagaa  ccagggggct  gccttggggc  2700
ctctggccag  agggaagcgg  gcagctctag  ccttgagat  tgtggtcaca  ttggggcttg  2760
tttaggattg  gagggccagg  tcacctcccc  agccacctc  ccttctctcc  tctggggctc  2820
ccactttagg  gcgactttgc  ccgagggcca  cgcattccat  cactccttta  gtgccttgaa  2880
tctcattcac  aagcagcccc  ctcccttccc  ctcccttct  cactctgttg  atgtaatcct  2940
cccaccccca  gtgtccatcc  taagacaggc  atcaaaagag  gccctaactt  tacttcccaa  3000
atggtgcttt  ttaaaaaaca  ccattcactac  attaggggca  attttttcac  acctccctgt  3060
cttcagaatg  taaagggtgg  ggattatgtc  tctgtttaat  atgcagcccc  ctgcactgtg  3120
gggtttgggg  catgttcagt  aataagaatg  aatacaatag  acaaaggggg  tgtgatgagt  3180
gttaacttgg  ttgttccaga  ggtaaaccag  gtctcaggga  agagcctgga  gctgctattg  3240
cttaggaagt  tgtacgtgcc  ttgaaatgtc  cactacctga  ggacgcagca  tctgggtggc  3300
ccggggctgc  tggggccaag  gagggacctt  gaccttctgg  tgcttgcccc  tccccggcct  3360
atgtgccatc  cgctgacgc  taggtcagcg  ctgccgggtc  ttaccaaata  gtgcttgctt  3420
ctcctaagtt  atttataaag  agaaatcact  aatggactct  actggtttga  gtgcttctga  3480
gctggatgac  cgaccgctg  tatgtttgtg  taattaattg  ccataataaa  ctttgatgag  3540
tccaaaaa

```

<210> 45

<211> 776

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7505082CB1

<400> 45

```

gcaaatcaac  ggcatccaga  aagccatgtc  ggactcggcg  ccagcgcccc  aagcgctaac  60
ccgctgaaag  tttctcagcg  aaatctcagg  gacgatctgg  accccgctga  gaggaactgc  120
ttttgagtga  gatgggtccc  gaggcctgga  ggagcggact  ggtgtatctc  accacggact  180
cccgcgcag  cgacccgctg  ctgaagaagc  ctggtgcagc  cagtccactg  gccagccgcc  240
agaacacgct  gcggagctgt  gatccgggtg  tctataggca  ggtgttgggt  gcagagagcg  300

```

```

ccccccccg acagcaagcc ccgcccaca cggactggcg tttctctcag gccagagac 360
ccggcaccag cggctcccaa aatggcgatg acaccggcac ctggcccaac aaccagtttg 420
acacagagat gctgcaagcc atgatccttg cgtccgccag tgaagctgct gatgggagct 480
ccaccctggg aggggggtgcc ggcaccatgg gattgagcgc ccgctacgga cccagttca 540
ccctgcagca cgtgcccagc taccgccaga atgtctacat cccaggcagc aatgccacac 600
tgaccaacgc agctggcaag cgggatggca agggcccagc aggtggcaat ggcaacaaga 660
agaagtcggg caagaaggag aagaagtaac atggaggcca ggccaagagc cacagggcgg 720
cctctcccca accagcccag cttctcctta cctgcaccca ggctcagag tttcag 776

```

<210> 46

<211> 2521

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7505139CB1

<400> 46

```

cgcccccgca cgccgaaaac aggggcctct cactgaccc ctgcgcgctc ccgcggggga 60
gcgtagtctc ggaggcggcg ccgcagggga ttgaggggtt gactgagcgt tgcgagcctt 120
agctttctcc ggaacgccag cgctgaggac acgatgtcgc ggctctcccg ctactgctt 180
tgggcccgca cctgcctggg cgtgctctgc gtgctgtccg cggacaagaa cagcaccag 240
caccggaacg tgacgacttt agcgcctatc tccaacgtaa cctcggcgcc ggtgacgtcc 300
ctcccgttgg tcaccactcc ggcaccagaa acctgtgaag gtcgaaacag ctgcgtttcc 360
tgttttaatg ttagcgttgt taatactacc tgcttttggg tagaatgtaa agatgagagc 420
tattgtttac ataactcaac agttagtgtg tgtcaagtgg ggaacacgac agacttctgt 480
tccgtttcca cggccactcc agtgccaaca gccaatctta caggtacaac aaataacact 540
gtgactccaa cctcacaacc tgtgcgaaag tctacctttg atgcagccag tttcattgga 600
ggaattgtcc tgggtcttga aataagatgc cacacaagga actacattcc agatttaaag 660
aaatgaaagg ataccattag tgtgtataac agattattgt tcatacttgt aaagcatctt 720
atgtcattga gaataataag aacagtgcct tagaagacag tgaaaggtaa gctctagctt 780
aatgtctatg atttgttctt tgacattaag gaaggtaagg attggtcaga ggatgtaact 840
tgatgtgagc agtagtaaac ctgttttact taccatactg ttaatatattt attgaaaatt 900
tatttcagag cggagaaaact taagctaaag tctgttatac agaattgaaa gccttcgtat 960
cttgaacctc ccaacatttt tcttatggct gttgaaaagt atagagctaa attgatttaa 1020
ttacactttc ctttgtactt taaaaaaaag tatgctagca ctattgtacc ttgaaaggat 1080
ttccaccaga ctgtcttgag tagtgacttc tttggtgagg caagaaggat atacattatt 1140
ttagaatcat ttactattta aatgagacaa tcatattatt ttagaatcat ttattttaaa 1200
tgagacaatc attttaagtt ttaagataac agaagtgacc aatgtaattt cacaacacct 1260
aaggattttt tgggtgatca ggttactgta gatttttact gattgtcctg gatgaataga 1320
ctgtgctttt tctttttctc tcccttcctt cttggtttcc catagtataa taagcatgca 1380
tactttaact tctatagttt tctcctttag agggctcgtc tcagtttttag aggtttactt 1440
ctcccttgcc tttgactcat tggactagtg cagaggcttt aagtagttta aaatgggctt 1500
ttgcttttct aggtcattaa cgttttttat ttagtttctt tagccaatag tggctgagtt 1560
tcgcacttga ttttcaatat tttatagtaa gaattgacaa actgcttttg ttcatttcat 1620
aaacaaactc tgcatttaga taactattaa aggttggttaa gatgaagatt tactgtttct 1680
ttgttactcg ttggtacagc tgtttgtttt acttgcacat ttgtacatat acttaatgtt 1740
ttcaagtgcc ttaattgttt aaaatctctg gcttcaaagt ttcttgggga aaggctcggtt 1800
tacctcacat tttttgtttc cattagtaat attctaggta cctcacaata tgtattatgg 1860
tgccatggct gttagttttt agtgagtgtc gtaggattaa ttcgaaaata ggcagaattc 1920
cattcctccc aagtgggcaa aaattagcta tactgatgta attgtcattt acctgggtat 1980
gaattccctg acacacattc atgtcaacat atgtagcaaa ttttgtgaaa acataacaat 2040
ttgaagcttc tgtaattttg agcactgctc taacaacaag cataatataa aattagtttag 2100
atthttgcaag tctacaaatg agctcttgca acagaactca cagccttttt acttttttcc 2160
cctaacttta gcaatgtagt atcttgagcc attaattttt gggttttttt taaaatccag 2220
aaggatatata gaaacctttt cagatttttc atctgatttg ttcttgacaga tgttcttcta 2280
tcaaatacct tattttacct tacagatatt tgttgcacag gcagatactg ctgtatttag 2340
acatttctat ttcagttcat taaaaactgc aaaaaccaat tgtatcatgt accaaactga 2400
cttaaaaataa atctacatgt ttattgaatt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2460
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaataacaa agaactgcag ccagggtctg 2520
g
2521

```

<210> 47
 <211> 1884
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7505234CB1

<400> 47
 agccgacgct cccttaccga gatactcagc taaagaagca gcaagcagga agaggaggct 60
 ttctaaggcg gtcgctccgg gaaatccggg ccctaggatt gtccactcat cccagtatca 120
 gcgagatacg gggagataga gtttagcgaca acgtgagcca gagctggagc acgtttggtg 180
 agagaccaga aagcaatgga ggccggagag gggaaggagc gcgttccgaa acaaaggcaa 240
 gtcctgatat tctttgtttt gctgggcata gctcaggcta gttgccagcc taggcactat 300
 tcagtggccg aggaacgga gactggctcc tttgtggcca atttgttaaa agacctggg 360
 ctggagatag gagaacttgc tgtgaggggg gccagggtcg tttccaaagg aaaaaaatg 420
 catttgcagt tcgataggca gaccggggat ttgtgtttaa atgagaaatt ggaccgggag 480
 gagctgtgcg gccccacaga gccctgtgtc ctacctttcc aggtgttact agaaaatccc 540
 ttgcagtttt ttcaggcgga gctacggatt agggacgtaa atgatcattc cccagttttc 600
 ctagacaaag aaatactttt gaaaattcca gaaagtatca ctcttggaac tacttttcta 660
 atagaacgtg cccaggactt ggatgtagga accaacagtc tccaaaatta cacaatcagt 720
 cccaatttcc actttcatct taatttacaa gacagtctcg atggcataat attaccacag 780
 ctggtgctga acagagccct ggatcgcgag gacgagcctg agatcagggtt aaccctcaca 840
 gcgctagatg gcgggagtc acccagggtcc ggcacggccc tggtagcgat tgaagtgtg 900
 gacatcaatg acaacgtccc agagtttgca aagctgctct atgaggtgca gatcccgag 960
 gacagccccg ttggatccca ggttgccatc gtctctgcca gggatttaga cattggaact 1020
 aatggagaaa tatcttatgc attttccaa gcatctgaag acattcgcaa aacgtttcga 1080
 ttaagtgcaa aatcgggaga actgctttta agacagaaac tggatttcga atccatccag 1140
 acatacacag taaatattca ggcgacagat ggtggggggc tatctggaac ttgtgtggta 1200
 tttgtccaag tgatggattt gaatgacaat cctccggaac taactatgtc gacacttatc 1260
 aatcagatcc cagaaaactt gcaggacacc ctcatgtctg tattcagcgt ttcagatcct 1320
 gactccggag acaacggaag gatggtgtgc tccatccaag atgatcttcc ttttttcttg 1380
 aaaccttctg ttgagaactt ttacactctg gtgataagca cggccctgga cgggagacc 1440
 agatccgaat acaacatcac catcacctgc accgacttgc ggacctgtc ccagagctac 1500
 cagtacgagg tgtgtctgac tggaggctcc gggacaaatg agttcaagtt cctgaagcca 1560
 attatcccca acttcgttgc tcagggtgca gagagggtta gcgaggcaaa tcccagtttc 1620
 aggaagagct ttgaattcac ttaagtgtta ataaggatct actgaggcta gtctcgttta 1680
 atttgtggaa agtccttttt tactgctttg cccattggag gtgtctcctt ttattagaaa 1740
 gtaaccatct taatccaat ctatgcatgt tactggatt tataaatgta tgagtttttt 1800
 tgcggtataa taaatgtaaa ttttctttgt attctaattg ttggttagtt tcattgcaat 1860
 ttaattgcat ttaaagtgtg aaag 1884

<210> 48
 <211> 1132
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7500227CB1

<400> 48
 cgcggaactg aacgcattgc ttcgggaccc aggacccct cgggcccgcac ccgccaggaa 60
 agactgaggg cgcggcctgc cccgcccggc tccctgcgcg gccgcccgcct cccgggacag 120
 aagatgtgct ccagggtccc tctgctgctg ccgctgctcc tgctactggc cctggggcct 180
 ggggtgacag gctgcccac cggctgccag tgcagccagc cacagacagt cttctgcact 240
 gcccgcaggg ggaccacggt gccccgagac gtgccaccgg acacggtggg gctgtacgtc 300
 tttgagaacg gcatcaccat gctcgacgca ggcagctttg cgggcctgcc gggcctgcag 360
 ctccctggac tgtcacagaa ccagatcgcc agcctgccc gcggggtctt ccagccactc 420
 atgggcttcc cagggcctgg cctccagtca cccctccaag caaagcccta catctaagcc 480
 agagagagac agggcagctg gggccgggct ctcagccagt gagatggcca gccccctcct 540
 gctgccacac cacgtaagtt ctcagtccca acctcgggga tgtgtgcaga cagggtgtg 600

```

tgaccacagc tgggccctgt tccctctgga cctcgggtctc ctcatctgtg agatgctgtg 660
gcccgactga cgaagccctaa cgtccccaga accgagtgcc tatgaggaca gtgtccgccc 720
tgccctccgc aacgtgcagt ccctgggcac ggccgggccc gcatgtgct ggtaacgcat 780
gcctgggccc tgctgggctc tcccactcca ggccggaccct gggggccagt gaagggaagt 840
cccggaaaga gcagagggag agcgggtagg cggctgtgtg actctagtct tggccccagg 900
aagcgaagga acaaaaagaaa ctggaaagga agatgcttta ggaacatgtt ttgctttttt 960
aaaatatata tatatttata agagatcctt tcccatttat tctgggaaga tgtttttcaa 1020
actcagagac aaggactttg gtttttgtaa gacaaacgat gatatgaagg ccttttgtaa 1080
gaaaaaataa aagatgaagt gtgaaaaaaa aaaaaaaaaa aaaaaaaaaa aa 1132

```

<210> 49

<211> 2391

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503676CB1

<400> 49

```

tgccctgccag ctagccggag ccgcgggtga gcgcggcgag cggcgaccct ggtgaggagc 60
gcggcgcgggg aggcacgttc cttagctccg ccgcggccgt cctccgcggc tcgaggactc 120
cgcttcccttc cctccccctcc cctgcgtccc ggccctggggg cttggcgcgcg ggagcggagg 180
gaagggacga aggaggagta ggtgaaagcg gggtgagggg cggaaggggt ccggcgcggg 240
gtgaggcgag ggctgcctct tgttctcccg ccgctgcgcg cgtctccttg tcgggtgccc 300
cggccagagg cgcgcggggc tgccgaggca cccgcactat gcaggcagac tgccggccgc 360
cgcatggcg agccgggccc tggtgagagc caggcgctgc ccgagtgct cccaagtccc 420
ggccgcggcc gccgcgcccg cctgggcccgc gctccccctc tcccgtccc tccctccctg 480
ctccaactcc tccctcctct ccatgcctct gttcctcctg ctcttacttg tccgtgctct 540
gctgctcgag gacgctggag cccagcaagg caaatactgt ggtctggggg tgcaaatgaa 600
ccattcaatt gaatcaaaag gcaatgaaat cacattgctg ttcattgagt gaatccatgt 660
ttctggacgc ggattttttg cctcatactc tgattatagat aaacaagatc taattacttg 720
tttggacact gcatccaaat ttttgaacc tgagttcagt aagtactgcc cagctggttg 780
tctgcttccc tttgctgaga tatctggaac aattcctcat ggatatagag attcctcgcc 840
attgtgcatg gctgggtgtgc atgcaggagt agtgtcaaac acgttggggc gccaaatcag 900
tgtttgtaatt agtaaaggta ttccctatta tgaaagtctt ttggctaaca acgtcacatc 960
tgtggtggga cacttatcta caagtctttt tacatttaag acaagtggat gttatggaac 1020
actgggggat gagtctgggt tgatcgcgga tcccaaaata acagcatcat ctgtgctgga 1080
gtggactgac cacacagggc aagagaacag ttggaaaccc aaaaaagcca ggctgaaaaa 1140
acctggaccg ccttgggctg cttttgccac tgatgaatac cagtggttac aaatagattt 1200
gaataaggaa aagaaaaata caggcattat aaccactgga tccaccatgg tggagcacia 1260
ttactatgtg tctgcctaca gaatcctgta cagtgatgat gggcagaaat ggactgtgta 1320
cagagagcct ggtgtggagc aagataagat atttcaagga acaaaagatt atcaccagga 1380
tgtgcgtaat aacttttttg caccaattat tgcacgtttt attagagtga atcctacca 1440
atggcagcag aaaattgcca tgaaaatgga gctgctcgga tgtcagttta ttcctaaagg 1500
tcgtcctcca aaacttactc aacctccacc tccctcggaac agcaatgacc tcaaaaacac 1560
tacagccctt ccaaaaatag ccaaaggctg tgccccaaaa ttacgcaac cactacaacc 1620
tcgcagtagc aatgaatttc ctgcacagac agaacaaaca actgccagtc ctgatatcag 1680
aaatactacc gtaactccaa atgtaaccaa agatgtagcg ctggctgcag ttcttgccc 1740
tgtgctgggt atggtcctca ctactctcat tctcatatta gtgtgtgctt ggcactggag 1800
aaacaggtta gtacataact agttcacctg agtccaaaac taccaaatgt gaagtagaag 1860
ctaaatatag aagatgaaaa tgtttacctg tttgagagtg agagttaagg taattattaa 1920
aatgaaaatt tcatgcttct cctttattcc cattaaaaat aaataagttc aattccacaa 1980
tcaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaacaaacag 2040
gaaaaaaaaa acaagaacaa gaaaaaaaaa aggagggggg gacaacaaag aagggaacag 2100
acgcaacaaa gaaaaaaaaa accgaaaacg aacagacgaa gagacttcac aaagacgcca 2160
ggaaagacat aaaaactgcc ggggcaacca aaagaaggag aaacaagaca aatacagcaa 2220
cagaaacgac agaagccaca aaccgcgcaa tagagagcaa ccggagcaaa aaaaaggaca 2280
cagatatcga gaacagcga gggcgaagaa cgggaagacg atagaagcga caagaagaaa 2340
gccacgcccc gacgatagaa gagcacaaca ggcgaagcca cgaggaaaaa c 2391

```

<210> 50

<211> 4550

<212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7503606CB1

<400> 50

```

atggggaagg agcaggagct ggtgcaggcg gtgaaggcgg aggacgtagg gaccgcgcag 60
aggctgctgc agaggccgcg gccgggaag gccacgcgta gcctccctgg gggccgcgg 120
agatggatgg atgggcgtgt ggaccagccg cgtgtgcggc tgcgcacgta tagccgtgtc 180
agtgtgtcag ggcacctgtg cgggcacgga cagggctctg cagagctcct gggttccacc 240
aagaagatca atgtcaactt ccaggaccgg gatggggttg ggtttggggg caagggtcag 300
ctcccagcat cccctcgccc cccaggcatg cggccgctgc actatgcggc ctggcagggc 360
cggaaggagc ccatgaagct ggtgctgaag gcgggctcgg ccgtgaacat cccgtctgat 420
gagggccaca tccccctgca cctggcgccc cagcatggtc actatgatgt gtctgagatg 480
ctgctacagc accagtctaa cccgtgcatg gtggacaact cggggaagac gcccctggac 540
ctggcctgcg agttcggccg cgttgggggt gtccagctgc tcctcagcag caatatgtgt 600
gcggcgctgc tggagccccg gccgggagac gccaccgacc ccaacggcac cagccctttg 660
cacctgcgag ctaaaaaacg ccacatcgac atcatcaggc tcctcctcca agccggcatc 720
gacattaacc gccagaccaa gtccggcagc gccctgcacg aggtgcgct ccggaagggc 780
acagaggtgg tgcggctgct gctggatagc gggatcaatg cccacgtgag gaacacctac 840
agccagacag cctgggacat cgtgcaccag ttcaccacgt cccaggccag cagggagatc 900
aagcagctgt tgcgaggttg gccaggcag ggcagagggg cagtggctgg ggctggggct 960
gggccagatc agcccgagg acttcggggg cccctcctg tatcccaga ggcctcagcg 1020
gccctgcagg tccgggcgac caaggattat tgcaacaatt acgacctgac cagcctcaac 1080
gtgaaggcag gggacatcat cacagtcctc gaggcagcat cggatggccg gtggaagggc 1140
tgcattccatg acaaccggac gggcaatgac cgggtgggct acttcccgtc ctccctgggc 1200
gaggccattg tcaagcgagc aggttcccga gcaggcactg aaccaagcct gccccaggga 1260
agcagctcat cgggaccctc tgcaccccc aaggagatct ggggtgctgag gaagcctttt 1320
gcaggtgggg accgaagcgg cagcattagc ggcattggct gcggccgggg cagcgggggt 1380
cacgcctac acgcgggctc tgaaggcgtc aagctcctgg caacggtgct tcccagaag 1440
tccgtctctg agtcggcccc gggggacagc ccgcgaagc ctccgaagg ctctgcagg 1500
gtggcccggt cccagcctcc agtggccac cgcgggcagg tctatgggga gcagccgcc 1560
aagaagctgg agccagcatc ggagggaag agctctgagg ccgtgagcca gtggctcacc 1620
gcgttccagc tgcagctcta cgcacccaac ttcacagcgc ccggctacga cctgcccacc 1680
atcagccgca tgactcccga ggacctcag gccattgggt tcaccaagcc gggccaccgg 1740
aagaagatcg cggcagagat cagcggccta agcatccctg actggctgcc tgagcacaaa 1800
cccgttaacc tggccgtgtg gctgtccatg atcggcctgg ccagtacta caaggtgttg 1860
gtggacaatg gctacgagaa cattgatttc acaccgaca tcacctggga ggacctgag 1920
gagatcggca tacccaagct ggggcaccag aagaagctga tgctcgctgt gaggaagctg 1980
gcagagctgc agaaggctga atacgccaag tatgaggggg gccccctgcg ccggaaggcg 2040
ccccagtctc ttgaagtgat ggccatcgag tcgcgcgcc cgcctgagcc cacaccggcc 2100
gactgccagt cccctaaaat gaccaccttc caggacagcg agctcagtga cgagctgcag 2160
gctgccatga ctggcccggc tgaggtgggg cccaccactg agaagccctc cagccacctg 2220
ccaccacccc cgagggccac cacgcggcag gactccagcc tgggtggctc ggcacggcac 2280
atgacagct cgcaggagct gctgggagat gggcccccct gcgccagcag ccccatgtct 2340
cgaagccagg agtacctcct ggatgagggc cccgcccccg gcaccccgcc cagggaggcc 2400
cggcccgccc gccacggcca cagcatcaag agggccagcg tgccccccgt gcctggcaag 2460
ccacggcagg tcctcccacc aggcactagc cacttcacgc cccccagac gccaccaaaa 2520
accgaccag gctctcccca ggcccttggg ggacctcatg gtccagcccc agctacggcc 2580
aaggtgaagc ccaccccgca gctgctgcgc ccgacgagc gccccatgtc acccgctcc 2640
ctgcccagct caccgacgca ccgcggcttt gacctagtc tgccccagcc cgtggagggc 2700
gaggtggggc cggtgcccc ggggcctgcg cccccaccg tgccgacggc tgtgcccaca 2760
ctgtgcctgc cccctgaggc cgacgcggag ccggggcggc ccaagaagcg ggccacagc 2820
ctgaatcgct atgcggcgct cgacagcgag ccggagcggg acgagctgct ggtgcctgcg 2880
gctgcgggcc cctatgccac ggtccagcgg cgcgtgggcc gcagccactc agtgagggcg 2940
ccggcaggtg ccgacaagaa cgtcaaccgc agccagtcct ttgcccgtgc gccccgaaag 3000
aaggggcccc cgcgcgcccc acccaagcgc tccagctcgg ccctggctag tgccaacctg 3060
gcggatgagc cggtgcctga cgccgagcct gaggatggcc tgctgggggt ccgggcacag 3120
tgccggcggg ccagtgcctt ggccggcagc gtggacacgg gtagtgccgg cagtgtgaag 3180
agcatcgcg ccatgctgga gctgtcctcc attgggggtg ggggcggggc tgcccgagg 3240
cctcctgagg gccaccccac tccccgcct gccagccag agccgggccc ggtggccacc 3300

```

| | | | | | | |
|-------------|------------|------------|------------|-------------|-------------|------|
| gtgctggcct | cagtgaacaa | caaagaggcc | atcgggcctg | gcggggaggt | ggtgaaccgg | 3360 |
| cgccgcacgc | tcagcggggc | agtcaccgga | cttctggcca | ctgcccgcgc | ggggcctggg | 3420 |
| gagtcggcag | accagggccc | ctttgtggag | gatggcactg | gccggcagcg | gcctcggggg | 3480 |
| ccctccaagg | gcgaggcggg | tgtcgaaggc | ccgcccttgg | ccaaggtgga | agccagcgcc | 3540 |
| acactcaaga | ggcgcacccg | ggccaagcag | aaccagcagg | agaacgtcaa | gttcacacctg | 3600 |
| accgagtctg | acacggtcaa | gcgcaggccc | aaggccaagg | agcgggaggg | cgggcctgag | 3660 |
| ccaccaccgc | cactgtccgt | gtaccataat | ggcactggca | ccgtgcgcgc | ccgaccggcc | 3720 |
| tcggagcagg | ctgggcctcc | ggagctgcct | ccaccgcccc | cgctgcgcga | acccccgccc | 3780 |
| accgactctg | cgcacctacc | cccattgccc | ccgcccagg | gcgaagcccg | gaagccggcc | 3840 |
| aagccgctg | tctctcccaa | gcccgtcctg | acgcagcctg | tgcccaagct | ccaggggtctg | 3900 |
| cccacaccca | cctccaagaa | ggtgccgctg | ccaggccctg | gcagcccaga | ggtgaagcgc | 3960 |
| gcccacggca | cgccaccgcc | cgtgtctccc | aagccgcgc | cgccgcccac | agcgcccaag | 4020 |
| cccgtaagg | cggtcgcggg | gctgccttcg | ggcagcgccg | gcccttcacc | cgcaccctcg | 4080 |
| cccgcgcgac | agccgccccg | cgccctcgcc | aagccgcccc | gtacgcgcgc | ctcgtctggc | 4140 |
| gccagccccg | ccaagccccc | gtcccccgcc | gcgcccgcgc | tgcacgtgcc | cgccaagccc | 4200 |
| ccgcgagccg | cgcgcgcgcg | cgcgcgcgcg | ccgcgcgcgc | cccccgcccc | gcccgaaagg | 4260 |
| gctcggccag | gggacagcgc | ccggcagaaa | ctggaggaga | caagcgcgctg | cctggccgcg | 4320 |
| gcgctgcagg | cggtggagga | gaagatccgg | caggaggacg | cgcagggccc | gcgcgactcg | 4380 |
| gcggcggaag | agagcactgg | cagcatcctg | gacgacatcg | gcagcatgtt | cgacgacctg | 4440 |
| gccgaccagc | tggatgccat | gctggagtga | acgccgcctg | gccggggcct | cccgcgcgcg | 4500 |
| ccgggcccctc | cccgcacact | gacctatacc | tcaggatggg | cgcgtctggg | | 4550 |

<210> 51

<211> 2727

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7500216CB1

<400> 51

| | | | | | | |
|------------|-------------|-------------|-------------|-------------|-------------|------|
| attctggggc | tcgggggagc | ccggacaccc | tctcagctcc | tgcccggggg | cccatgtagt | 60 |
| cccttctgcc | ctgtgcctcg | gtgcctgtga | cctgagcccc | ttgggtgacc | ctgcactcgt | 120 |
| ccaacttggg | ccaaacgact | gcccctcctt | ctggcagtg | gctggaccag | ccggccagcg | 180 |
| ggagccccct | tggcagaagc | cggtcgtaaa | ggatcataaa | ctggcgccgt | ctggctgggg | 240 |
| cgaaggtcgc | tgaggttaga | actgcgccag | tcctagacgc | cagacccgct | cagaccctcc | 300 |
| tgccaggtga | cagccgccaa | gatggggctc | tgggcccctg | tgtggcctcc | cctgctgttc | 360 |
| accgggctgc | tcgtccgacc | cccggggacc | atggcccagg | cccagtagct | ctctgtgaac | 420 |
| aaggacatct | ttgaagtaga | ggagaacaca | aatgtcaccg | agccgctggg | ggacatccac | 480 |
| gtcccggagg | gccaggagg | gaccctcgga | gccttggtcca | ccccctttgc | atttcggatc | 540 |
| cagggaaccc | agctgtttct | caacgtgact | cctgattacg | aggagaagtc | actgcttgag | 600 |
| gctcagctgc | tgtgtcagag | cggaggcaca | ttgggtgacc | agctaagggt | gttcgtgtca | 660 |
| gtgctggacg | tcaatgacaa | tgcccccgaa | ttccccttta | agaccaagga | gatattgggtg | 720 |
| gaggaggaca | cgaagtgaa | ctccaccgtc | atccctgaga | cgcaactgca | ggctgaggac | 780 |
| cgcgacaagg | acgacattct | gttctacacc | ctccaggaaa | tgacagcagg | tgccagtgac | 840 |
| tacttctccc | tggtgagtgt | aaaccgtccc | gccctgaggc | tggaccggcc | cctggacttc | 900 |
| tacgagcggc | cgaacatgac | cttctggctg | ctgggtgcggg | acactccggg | ggagaatgtg | 960 |
| gaaccagccc | acactgccac | cgccacacta | gtgctgaacg | tggtgcccgc | cgacctgcgg | 1020 |
| cccccggtgg | tcctgcccctg | caccttctca | gatggctacg | tctgcattca | agctcagtag | 1080 |
| cacggggctg | tccccacggg | gcacatactg | ccatctcccc | tcgtcctgcg | tccccgaccc | 1140 |
| atctacgtcg | aggacggaga | ccgcggcatc | aaccagcccc | tcactctacag | catcttttag | 1200 |
| ggaaacgtga | atggtacatt | catcatccac | ccagactcgg | gcaacctcac | cgtggccagg | 1260 |
| agtgtcccca | gccccatgac | cttcttctctg | ctgggtgaagg | gccaacaggc | cgaccttgcc | 1320 |
| cgctactcag | tgaccacagg | caccgtggag | gctgtggctg | cggccggggg | cccgcgccgc | 1380 |
| ttcccccaga | gactgtatcg | tggcaccgtg | gcgcgtggcg | ctggagcggg | cgttgtgtgc | 1440 |
| aaggatgcag | ctgccccttc | tcagcctctg | aggatccagg | ctcaggaccc | ggagtctctg | 1500 |
| gacctcaact | cggccatcac | atatcgaatt | accaaccact | cacacttccg | gatggaggga | 1560 |
| gaggttgtgc | tgaccaccac | cacactggca | caggcggggg | ccttctacgc | agaggttag | 1620 |
| gcccacaaca | cggtgacctc | tggcaccgca | accacagtca | ttgagataca | agtttccgaa | 1680 |
| caggagcccc | cctccacagc | acagacccca | gaggcaggaa | cctctcagcc | gatgcccccc | 1740 |
| ggtatgggaa | ccagcacctc | ccaccaacca | accacaccgc | gtggggggcac | agcacagacc | 1800 |
| ccagagccag | gaacctctca | gccgatgccc | ctcagcaaga | gcaccccatc | ttcagggtggc | 1860 |

| | | | | | | |
|-------------|-------------|-------------|------------|--------------|-------------|------|
| ggccccctcgg | aggacaagcg | cttctcgggtg | gtggatatgg | cggccccctggg | cgggggtgctg | 1920 |
| ggtgcgctgc | tgctgctggc | tctccttggc | ctcgccgtcc | ttgtccacaa | gcactatggc | 1980 |
| ccccggctca | agtgcctgctg | tggcaaaagct | ccggagcccc | agcccccaagg | ctttgacaac | 2040 |
| caggcggttcc | tccctgacca | caaggccaac | tgggcgcccc | tccccagccc | cacgcacgac | 2100 |
| cccaagcccc | cggaggcacc | gatgcccgc | gagcccgcc | cccccgcccc | tgcctcccca | 2160 |
| ggcggtgccc | ctgagcccc | cgcagcggcc | cgagctggcg | gaagccccac | ggcggtgagg | 2220 |
| tccatcctga | ccaaggagcg | gcgccagag | ggcggttaca | aggctgtctg | gtttggcgag | 2280 |
| gacatcgga | cggaggcaga | cgtggctgtt | ctcaacgcgc | ccaccctgga | cgtggatggc | 2340 |
| gccagtgaact | ccggcagcgg | cgatgagggc | gagggcgccg | ggaggggtgg | gggtccctac | 2400 |
| gatgcgcccc | gtgggtgatga | ctcctacatc | taagtggccc | ctccaccctc | tccccagcc | 2460 |
| gcacgggcac | tggaggtctc | gctccccag | cctccgaccc | gaggcagaat | aaagcaaggc | 2520 |
| tcccgaaaacc | caggccatgg | cgtggggcag | gcgcgcgggt | ccatgggggt | ccattcact | 2580 |
| cagtcccctg | tcgtcattag | cgcttgagcc | caggtgtgca | gatgaggcgg | tgggtctggc | 2640 |
| cacgctgtcc | ccacccaag | gctgcagcac | ttcccgtaaa | ccacctgcag | tgcccgcgcg | 2700 |
| cttcccagag | ctctgtgcca | gctagtc | | | | 2727 |

<210> 52

<211> 4013

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7099880CB1

<400> 52

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ggataggaga | gggagaagga | gagaggggca | gaggggaagg | gagagagggga | gagcacgcga | 60 |
| gacggaaagg | agcgctcag | agtctctgaa | gcacgcaaga | gataaccgat | taggaatttt | 120 |
| tcgggcaact | gtcaccgcga | tagctgtcag | agaatcatca | tcaccgcaac | tctgacgttt | 180 |
| cctacaagaa | gttagagact | taagcagtat | tggcatcgga | tggaaatggg | atgcatgctg | 240 |
| ctgtggaaaa | tatccctgag | ctgaagaagt | gcaactatgt | gttgtgtgtg | ctaaatgccc | 300 |
| agaagaacct | agaccagtg | ggtaagagag | acgaaatgtg | gagagaaatga | ctaacgacaa | 360 |
| atccgtgaat | ttgttctctg | gagttgctaa | tttgccagcc | cgaagcaaca | cattttacaa | 420 |
| ggaggaaaca | ttttgctgct | tctgatttct | ccccaaactg | gtgctaccag | atgcatggct | 480 |
| ttctatgtgt | tggaaacaaat | cattcctaac | tgcagcatac | tgggaaaggg | gaaagggttg | 540 |
| ggcaactaac | cggctgtctc | caaataagtg | atgagatata | atgcatcctc | cagtccatgc | 600 |
| acgcttcctc | ttgcaatttg | tgcattcatc | ctcagtgga | acctttaaac | cctaaaatcc | 660 |
| aggaaaagaa | aataaataca | ttatcatgga | cctgagggat | ttttacctgt | tggtctgctc | 720 |
| gattgcctgt | ttaaggctgg | attccgcaat | agctcaagaa | cttattttaca | ctattagaga | 780 |
| ggaattgcct | gaaaaatgtgc | ccataggaaa | cataccaaag | gatctgaaca | tttctcacat | 840 |
| caatgctgcc | acagggacca | gcgcagcct | tgtctacaga | ctggttttcta | aagctgggga | 900 |
| tgcccctttg | gtgaaagtat | ccagcagcac | tggggaaatt | ttcacaacct | ccaacagaa | 960 |
| agacagagaa | aaactctgtg | ctggcgccctc | atatgctgag | gagaatgagt | gtttctttga | 1020 |
| acttgagggt | gtgatcctcc | ccaatgattt | cttcaggctg | atcaaaataa | aaataattgt | 1080 |
| caaggatacc | aatgataatg | cccccatgtt | tccatctcct | gtcatcaata | tttccattcc | 1140 |
| agaaaacact | ttgatcaaca | gccgctttcc | aattccatca | gcaacagatc | ctgacacagg | 1200 |
| cttcaatggg | gtacagcatt | atgaattgtt | aaatgggcag | agtgtttttg | gactgggat | 1260 |
| cgtggaaact | ccagagggag | agaagtggcc | acaactgatt | gttcagcaaa | acttgggat | 1320 |
| agaacagaaa | gatacctatg | tgatgaaaat | caaagtagag | gatggaggca | ctccacagaa | 1380 |
| atccagtacg | gccatactgc | aggtcacagt | aagtgatgta | aatgacaaca | ggccagtgtt | 1440 |
| taaagagggt | ccaagtggag | tgcataattcc | agagaatgct | cccgtaggta | cctctgtaat | 1500 |
| tcagctccat | gccactgatg | cagatatagg | cagtaatgct | gaaatccggg | acattttttg | 1560 |
| tgcccagggtc | gcccctgcaa | ccaaaagact | ctttgcttta | aataatacta | ctgggctgat | 1620 |
| tacagttcag | aggctccttag | atagagagga | gacagccatt | cacaaagtga | cagtgcctggc | 1680 |
| tagtgacggc | agctccactc | ctgctcgagc | aacgggtacc | atcaatgtca | ccgatgtaaa | 1740 |
| tgataaccct | cctaataatag | acctcaggta | cattataagt | cccatcaatg | gcaccgtgta | 1800 |
| tttatctgag | aaagatcctg | tcaatacaaa | gattgccccta | attacagttt | cagataagga | 1860 |
| cacagatgtg | aatggcaaaag | tgatctgttt | tattgaaaga | gaggtcccat | ttcatttgaa | 1920 |
| ggcggtatat | gacaaccaat | atttgtttaga | gacctcttct | ttgttggaact | atgagggcac | 1980 |
| caaagaattc | agcttttaaaa | ttgttgccctc | tgattctggg | aagcccagtt | taaatcagac | 2040 |
| tgccctggta | aggggttaagc | ttgaggatga | aaatgacaac | ccaccaattt | tcaaccagcc | 2100 |
| tgtaatttgag | ctgtcagttt | ctgaaaacaa | ccgacgtggg | ttataacttaa | caactattag | 2160 |
| tgccacagat | gaagacagtg | ggaaaaatgc | agacattgtt | tatcagcttg | gaccgaatgc | 2220 |


```

ctccttctttt gatctggacc gaaaaacagg agtttttgaca gcctccagag tattttgacag 2280
agaagaacaa gaacgattca tttttacagt aactgccagg gacaatggga cccctcccct 2340
ccaaagccaa ggggctgtga ttgttactgt tctggatgag aatgacaata gcccgaagtt 2400
tactcataat ctttttcaat tttttgtgtc tgagaatctg ccaaagtata gtactgtggg 2460
ggtaatcaca gtgacagatg cagatgctgg agagaataaa gctgtgactc tttccattct 2520
aaatgacaat gataattttg tgttgatcc ctattctgga gtcataaagt caaatgtctc 2580
atttgataga gagcagcaga gttcctacac ttttgatgtc aaagccactg atggaggaca 2640
accacctcgt tctctactgt caaaagtaac tatcaacgtc atggatgtca atgacaacag 2700
cccagttgtc atttctccac cgtctaatac ttcttttaag ttgggtgccc tctcagccat 2760
tcttggtctc gtggtagcag aagtttttgc agtggatgtt gacactggaa tgaacgctga 2820
actaaagtat actatagtga gtggaaacaa taaaggctta ttccggattg atccagtaac 2880
aggtaacatt actctggaag aaaaaccagg acctactgat gtgggattgc atcgtttggg 2940
ggtcaacata agtgacctgg ggtaccctaa gtctttgcac acgcttgtgc ttgtattcct 3000
ttatgttaac gacactgctg gaaatgcctc ctatatctat gacttgatcc gcaggactat 3060
ggagaccccg ttggacagga acatagggga tagtagccaa ccctatcaaa atgaggacta 3120
tctaaccatc atgattgcc aatcgccgg tgccatgggtg gtcattgttg tgatcttcgt 3180
caccgttctg gtgcgctgtc gccatgcac aaggttcaaa gcagctcaga ggagcaagca 3240
aggtgccgaa tggatgtccc caaaccagga gaacaagcaa aacaagaaaa agaaaagaaa 3300
gaaaaggaag tctcccaaaa gctctctttt gaactttgtt actatcgaag agtccaaacc 3360
cgatgatgca gttcatgaac ctatcaatgg gacaataagc ctgccggctg aactggagga 3420
gcaaagtata ggaagatttg actggggccc ggcacctcca acaacattca agcctaacag 3480
tcttgacctg gccaagcact acaaatctgc ttctccacag cctgcttttc atctcaaacc 3540
agacactcca gtttccgtga aaaagcacca cgtgattcag gaactccctt tggacaacac 3600
ctttgttggg ggttgtgaca ccctttctaa acgctcttcc actagtccag atcacttcag 3660
tgacctcagag tgcagttccc aaggaggctt caagacaaag ggccccttac acaccagaca 3720
ggtaaaccgag cacttttact ggtctataag tactgcatac aagtgccag tcaaccagta 3780
ttaacgtgcc agtatgtcta ttgttttggg ctaactttag cttagttaga aagaggtaaa 3840
aaaaaacttg accccttttc tttctctctg gggctcagtc aagaattttt agaacagctt 3900
tgaaatttct gagttctgag aatattttct ctttctcagtt tccttcaaat ggcaaaagat 3960
gacaagcaga aattttcaaa tgtccagata tgttgtgaag agaaaaactg tct 4013

```

<210> 53

<211> 2768

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 871513CB1

<400> 53

```

cgaaggaggg agccagcgcg cccagcccag cccaggctcg gacagagggg gcggggagag 60
gtggagcgcg gggagccagg cgagggggcc cgacgacggg actccattag ccgctccggc 120
cacaggcagc gcttcgccag ccgaggaacc ggacgcggag accgccgccc cgcgagcctc 180
cagccctcgt cctgttgccg cgcgagtcct gggcccgagg cgctaggagc gcgcggaagg 240
agccatggct ctggacggga taaggatgcc aatggctgc tactcgagac ggacgtggga 300
actgagtgtc catgtgacgg acctgaaccg cgatgtcacc ctgagagtga ccggcgaggt 360
gcacattgga ggcgtgatgc ttaagctggg ggagaaactc gatgtaaaaa aagattgggtc 420
tgacctatgt ctctgggtggg aaaagaagag aacttggctt ctgaagacac attggacctt 480
agataagtat ggtatttcagg cagatgctaa gcttcagttc acccctcagc acaaactgct 540
ccgcttcgag cttcccaaca tgaagtatgt gaaggtgaaa gtgaatttct ctgatagagt 600
cttcaaagct gtttctgaca tctgtaagac ttttaataac agacaccccg aagaactttc 660
tctcttaag aaaccagag atccaacaaa gaaaaaaaag aagaagctag atgaccagtc 720
tgaagatgag gcacttgaat tagagggggc tcttatcact cctggatcag gaagtatata 780
ttcaagccca ggactgtata gtaaaacaat gaccccccact tatgatgctc atgatggag 840
ccccttgtca ccaacttctg cttggtttgg tgacagtgtt ttgtcagaag gcaatcctgg 900
tatacttgtc gtcagtcaac caatcacgtc accagaaatc ttggcaaaaa tgttcaagcc 960
tcaagctctt cttgataaag caaaaatcaa ccaaggatgg cttgattcct caagatctct 1020
catggaacaa gatgtgaagg aaaaatgagg cttgtgtctc cgattcaagt attacagctt 1080
ttttgatttg aatccaaagt atgatgcaat cagaatcaat cagcttttatg agcaggccaa 1140
atgggcccatt ctctggaag agattgaatg cacagaagaa gaaatgatga tgtttgcagc 1200
cctgcagtat catatcaata agctgtcaat catgacatca gagaatcatt tgaacaacag 1260
tgacaaagaa gttgatgaag ttgatgtctc ctttctcagac ctggagatta ctctggaagg 1320

```

```

gggtaaaacg tcaacaattt tgggtgacat tacttccatt cctgaacttg ctgactacat 1380
taaagttttc aagccaaaaa agctgactct gaaagggttac aaacaatat ggtgcacett 1440
caaagacaca tccattttctt gttataagag caaagaagaa tccagtggca caccagctca 1500
tcagatgaac ctccaggggat gtgaagttac cccagatgta aacattttcag gccaaaaatt 1560
taacattaaa ctcttgattc cagttgcaga aggcattgaat gaaatctggc ttcgttgtga 1620
caatgaaaaa cagtatgcac actggatggc agcctgcaga ttagcctcca aaggcaagac 1680
catggcggac agttcttaca acttagaagt tcagaatat ctttctcttc tgaagatgca 1740
gcatttaaac ccagatcctc agttaatacc agagcagatc acgactgata taactcctga 1800
atgtttgggtg tctccccgct atctaaaaaa gtataagaac aagcagccag gctatataag 1860
agatttgata acagcgagaa tcttggaggc ccatcagaat gtagctcaga tgaagtcta 1920
tgaagccaag atgagattta ttcaagcttg gcagtcacta cctgaatttg gcatcactca 1980
cttcattgca aggttccaag ggggcaaaaa agaagaactt attggaattg catacaacag 2040
actgattcgg atggatgcca gcaactggaga tgcaattaaa acatggcgtt tcagcaacat 2100
gaaacagtgg aatgtcaact gggaaatcaa aatgggtcacc gtagagtttg cagatgaagt 2160
acgattgtcc ttcatattgta ctgaagtaga ttgcaaagtg gttcatgaat tcattggtgg 2220
ctacatatatt tctcacaacac gtgcaaaaga ccaaacgag agtttagatg aagagatgtt 2280
ctacaaactt accagtgggt ggggtgtgaat agaaatactg tttaatgaaa ctccacggcc 2340
ataacaatat ttaactttta aagctgtttg ttatatgctg cttataaag taagcttgaa 2400
atztatcatt ttatcatgaa aacttctttg ccttaccaga ccagttaata tgtgactaa 2460
acaagcacga ctattaatct atcatgttat gatataataa acttgaattt ggcacacatt 2520
ccttagggcc atgaattgaa aactgaaata gtgggcaaat caggaacaaa ccactagta 2580
tttactgatt taagctagcc aaactgtaag aaacaagcca tctattttta agctatccag 2640
ggcttaacct atatgaactc tatttatcat gtctaattgca tgtgatttaa tgtatgttta 2700
atgtgatatc atgtttttaa atatcctact tctggtagcc atttaattcc tccccctacc 2760
cccaaaaa

```

<210> 54
 <211> 5738
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 8057640CB1

```

<400> 54
gggaagagga tggcagggcc acgccccagc ccatggggcca ggctgctcct ggcagccttg 60
atcagcgtca gcctctcttg gaccttggca aaccgctgca agaaggcccc agtgaagagc 120
tgcacggagt gtgtccgtgt ggataaggac tgcgcctact gcacagacga gatgttcagg 180
gaccggcgct ccaacaccca ggcggagctg ctggcgcgcg gctgccagcg ggagagcatc 240
gtggtcatgg agagcagctt ccaaatacga gaggagacc agattgacac caccctgcgg 300
cgcagccaga tgtcccccca aggcctgcgg gtccgtctgc ggcccggtga ggagcggcat 360
tttgagctgg aggtgtttga gccactggag agccccgtgg acctgtacat cctcatggac 420
ttctccaact ccatgtccga tgatctggac aacctcaaga agatggggca gaacctggct 480
cgggtcctga gccagctcac cagcgactac actattggat ttggcaagtt tgtggacaaa 540
gtcagcgtcc cgcagacgga catgaggcct gagaaactga aggagccttg gcccaacagt 600
gacccccctt tctccttcaa gaacgtcatc agcctgacag aagatgtgga tgaagtccgg 660
aataaaactgc agggagagcg gatctcaggg aacctggatg ctcttgaggg cggcttcgat 720
gccatcctgc agacagctgt gtgcacgagg gacattggct ggcgcccgga cagcaccac 780
ctgctggtct tctccaccga gtcagccttc cactatgagg ctgatggcgc caacgtgctg 840
gctggcatca tgagccgcaa cgatgaacgg tgccacctgg acaccacggg cacctacacc 900
cagtcacaga cacaggacta cccgtcgggt gccacctgg tgcgcctgct cgccaagcac 960
aacatcatcc ccatctttgc tgtcaccaac tactoctata gctactacga gaagcttcac 1020
acctatttcc ctgtctcttc actgggggtg ctgcaggagg actcgtccaa catcgtggag 1080
ctgctggagg aggccttcaa tcggatccgc tccaacctgg acatccgggc cctagacagc 1140
ccccgaggcc ttcggacaga ggtcacctcc aagatgttcc agaagacgag gactgggtcc 1200
tttcacatcc ggcgggggga agtgggtata taccaggtgc agctgcgggc ccttgagcac 1260
gtggatggga cgcacgtgtg ccagctgccg gaggaccaga agggcaacat ccactgtaa 1320
ccttcttctt ccgacggcct caagatggac cggggcatca tctgtgatgt gtgcacctgc 1380
gagctgcaaa aagaggtgcg gtcagctcgc tgcagcttca acggagactt cgtgtgcgga 1440
cagtgtgtgt gcagcgaggg ctggagtggc cagacctgca actgctccac cggctctctg 1500
agtgcatttc agccctgcct gcgggagggc gaggacaagc cgtgctccgg ccgtggggag 1560
tgccagtgcg ggcactgtgt gtgctacggc gaaggccgct acgaggttca gttctgcgag 1620

```

| | | | | | | |
|-------------|-------------|------------|-------------|-------------|------------|------|
| tatgacaact | tccagtgtcc | ccgcacttcc | gggttccctc | gcaatgaccg | aggacgctgc | 1680 |
| tccatggggc | agtgtgtgtg | tgagcctggg | tggaacaggc | caagctgtga | ctgtcccttc | 1740 |
| agcaatgcca | cctgcacoga | cagcaatggg | ggcatctgta | atggacgtgg | ccactgtgag | 1800 |
| tgtggccgct | gccactgcca | ccagcagtcg | ctctacacgg | acaccatctg | cgagatcaac | 1860 |
| tactcggcga | tccacccggg | cctctgcgag | gacctacgct | cctgcgtgca | gtgccaggcg | 1920 |
| tggggcaccg | gcgagaagaa | ggggcgacag | tgtgaggaat | gcaacttcaa | ggtcaagatg | 1980 |
| gtggacgagc | ttaagagagc | cgaggaggtg | gtgggtgcgt | gctccttccg | ggacgaggat | 2040 |
| gacgactgca | cctacagcta | caccatggaa | ggtgacggcg | cccctggggc | caacagcact | 2100 |
| gtcctgggtg | acaagaagaa | ggactgccct | ccgggtccct | tctggtggct | catccccctg | 2160 |
| ctcctccctc | tcctgccgct | cctggccctg | ctactgctgc | tatgctggaa | gtactgtgcc | 2220 |
| tgctgcaagg | cctgcctggc | acttctcccg | tgctgcaacc | gaggtcacat | ggtgggcttt | 2280 |
| aaggaagacc | actacatgct | gcgggagaa | ctgatggcct | ctgaccactt | ggacacgccc | 2340 |
| atgctgcgca | gcgggaacct | caagggccgt | gacgtggtcc | gctggaaggt | caccaacaac | 2400 |
| atgcagcggc | ctggcttttg | cactcatgcc | gccagcatca | acccacacaga | gctggtgccc | 2460 |
| tacgggtgtg | ccttgcgcct | ggcccgccct | tgacacgaga | acctgctgaa | gcctgacact | 2520 |
| cgggagtcg | cccagctgcg | ccaggaggtg | gaggagaacc | tgaacgaggt | ctacaggcag | 2580 |
| atctccggtg | tacacaagct | ccagcagacc | aagtctccgg | agcagcccaa | tgccgggaaa | 2640 |
| aagcaagacc | acaccattgt | ggacacagtg | ctgatggcgc | cccgtctggc | caagccggcc | 2700 |
| ctgctgaagc | ttacagagaa | gcaggtggaa | cagaggccct | tccacgacct | caaggtggcc | 2760 |
| ccgggctact | acaccctcac | tgcagaccag | gacgcccggg | gcatggtgga | gttccaggag | 2820 |
| ggcgtggagc | tggtggacgt | acgggtgccc | ctctttatcc | ggcctgagga | tgacgacgag | 2880 |
| aagcagctgc | tggtggaggg | catcgacgtg | cccgaggcca | ctgccaccct | cggccgcgcg | 2940 |
| ctggtaaaaca | tcaccatcat | caaggagcaa | gccagagacg | tggtgtcctt | tgagcagcct | 3000 |
| gagttctcgg | tcagccgcgg | ggaccaggtg | gcccgcaccc | ctgtcatccg | gcgtgtcctg | 3060 |
| gacggcggga | agtcccaggt | ctcctaccgc | acacaggatg | gcaccgcgca | gggcaaccgg | 3120 |
| gactacatcc | ccgtggaggg | tgagctgctg | ttccagcctg | gggaggcctg | gaaagagctg | 3180 |
| cagtgtaagc | tccgtgagct | gcaagaagtt | gactccctcc | tgccggggccg | ccaggtcccg | 3240 |
| cgtttccacg | tcctagctcag | caaccctaag | tttggggccc | acctggggcca | gccccactcc | 3300 |
| accaccatca | tcatcaggga | cccagatgaa | ctggaccgga | gcttcacgag | tcagatgttg | 3360 |
| tcatcacagc | cacccccctca | cggcgacctg | ggcgccccgc | agaaccccaa | tgctaaggcc | 3420 |
| gctgggtcca | ggaagatcca | tttcaactgg | ctgccccctt | ctggcaagcc | aatggggtac | 3480 |
| agggtaaaagt | actggattca | gggtgactcc | gaatccgaag | cccacctgct | cgacagcaag | 3540 |
| gtgcccctcag | tggagctcac | caacctgtac | ccgtatttgcg | actatgagat | gaaggtgtgc | 3600 |
| gcctacgggg | ctcaggggcga | gggaccttac | agctccctgg | tgctctgccc | cacccaccag | 3660 |
| gaagtgccca | gcgagccagg | gcgtctggcc | ttcaatgtcg | tctcctccac | ggtgaccacg | 3720 |
| ctgagctggg | ctgagccggc | tgagaccaac | ggtgagatca | cagcctacga | ggtctgctat | 3780 |
| ggcctggtca | acgatgacaa | ccgacctatt | gggcccata | agaaagtgt | ggttgacaac | 3840 |
| cctaagaacc | ggatgctgct | tattgagaac | cttcggggagt | cccagcccta | ccgctacacg | 3900 |
| gtgaaggcgc | gcaacggggc | cggctggggg | cctgagcggg | aggccatcat | caacctggcc | 3960 |
| acccagccca | agaggcccat | gtccatcccc | atcatccctg | acatccctat | cgtggagccc | 4020 |
| cagagcgggg | aggactacga | cagcttccct | atgtacagcg | atgacgttct | acgctctcca | 4080 |
| tcgggcagcc | agaggcccag | cgtctccgat | gacactgagc | acctggtgaa | tggccgggat | 4140 |
| gactttgcct | tcccgggcag | caccaactcc | ctgcacagga | tgaccacgac | cagtgtctgt | 4200 |
| gcctatggca | cccacctgag | cccacacgtg | ccccaccgcg | tgctaagcac | atcctccacc | 4260 |
| ctcacacggg | actacaactc | actgaccgcg | tcagaacact | cacactcgac | cacactgccc | 4320 |
| agggactact | ccacctcac | ctccgtctcc | tcccagcgcc | tccctcccat | ctgggaacac | 4380 |
| gggaggagca | ggcttccgct | gtcctggggc | ctggggctcc | ggagtccggc | tcagatgaaa | 4440 |
| gggttccccc | cttccagggg | cccacgagac | tctataatcc | tggtctgggag | gccagcagcg | 4500 |
| ccctcctggg | gcccagactc | tcgcctgact | gctggtgtgc | ccgacacgcc | caccgcctg | 4560 |
| gtgttctctg | ccctggggcc | cacatctctc | agagttagct | ggcaggagcc | gcggtgcgag | 4620 |
| cggccgctgc | agggctacag | tgtggagtac | cagctgctga | acggcgggtg | gctgcacg | 4680 |
| ctcaacatcc | ccaaccctgc | ccagacctgc | gtgggtgggg | aagacctcct | gcccacacc | 4740 |
| tcctacgtgt | tccgcgtgcg | ggcccagagc | caggaaggct | ggggccgaga | gcgtgagggt | 4800 |
| gtcatcacca | ttgaatccca | ggtgcacccg | cagagcccac | tgtgtccctt | gccaggctcc | 4860 |
| gccttcaact | tgagcactcc | cagtgcacca | ggcccgcctg | tgcttactgc | cctgagccca | 4920 |
| gactcgctgc | agctgagctg | ggagcggcca | cggaggccca | atggggatat | cgctcggtac | 4980 |
| ctggtgacct | gtgagatggc | ccaaggagga | gggcccagcca | ccgcattccg | ggtggatgga | 5040 |
| gagagccccg | agagccggct | gaccgtgccc | ggcctcagcg | agaacgtgcc | ctacaagttc | 5100 |
| aaggtgcagg | ccaggaccac | tgagggtctc | ggccagagag | gcgagggcat | catcccata | 5160 |
| gagtcacagg | atggaggacc | cttcccgcag | ctgggcagcc | gtgccgggct | cttccagcac | 5220 |
| ccgctgcaaa | gcgagtacag | cagcatcacc | accacccaca | ccagcggcac | cgagcccttc | 5280 |
| ctagtggatg | ggctgaccct | gggggcccag | cacctggaga | caggcggctc | cctcaccg | 5340 |
| catgtgaccc | aggagtttgt | gagccggaca | ctgaccacca | gcggaacctt | tagcaccac | 5400 |

| | | | | | | |
|------------|-------------|------------|-------------|------------|------------|------|
| atggaccaac | agttcttcca | aacttgaccg | caccctgccc | cacccccgcc | acgtcccact | 5460 |
| aggcgtcctc | ccgactcctc | tcccggagcc | tcctcagcta | ctccatgcct | gcacccctgg | 5520 |
| gggcccagcc | cacccgcatg | cacagagcag | gggctaggtg | tctcctggga | ggcatgaagg | 5580 |
| gggcaaggtc | cgtcctctgt | gggcccgaac | tattttgtaac | caagagctgg | gagcagcaca | 5640 |
| aggacccagc | ctttgtttctg | cacttaataa | atgtttgcta | ctgaaacgaa | gtgcacacac | 5700 |
| tgcgcttcta | acggggccagt | cgccctaagc | ggacgggc | | | 5738 |

<210> 55

<211> 1551

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7505913CB1

<400> 55

| | | | | | | |
|------------|-------------|-------------|-------------|-------------|------------|------|
| gaggacgggg | aggacgacag | cgcccaactgt | ggattggaca | gtgtcaaaaa | gagggggcgg | 60 |
| ccctactgaa | ggggcggttg | ggcgacgaag | ggaagagtct | tttcagcgct | gaggactggc | 120 |
| gctgaggagg | cggcgggtgg | tcccggggcg | tttgagcggg | ctcacccgag | cccgcggggc | 180 |
| aacgcggatc | caggccccgac | tggcgggacc | gccccggatt | ccccgcgggc | cttcctagcc | 240 |
| gccatggagg | acggcgctcta | tgaacccccca | gacctgactc | cggaggagcg | gatggagctg | 300 |
| gagaacatcc | ggcggcgga | gcaggagctg | ctgggtggaga | ttcagcgctc | gcgggaggag | 360 |
| ctcagtgaag | ccatgagcga | ggtggagggg | ctggaggcca | atgagggcag | taagaccttg | 420 |
| caacggaacc | ggaagatggc | aatgggcagg | agaagttca | acatggaccc | caagaagggg | 480 |
| atccagttct | tgggtggagaa | tgaactgctg | cagaacacac | ccgaggagat | cgcccgcttc | 540 |
| ctgtacaagg | gcgaggggct | gaacaagaca | gccatcgggg | actacctggg | ggagagggaa | 600 |
| gaactgaacc | tggcagtgc | ccatgctttt | gtggatctgc | atgagttcac | cgacctcaat | 660 |
| ctggtgcagg | ccctcaggca | gtttctatgg | agctttcgcc | tacccgagga | ggcccagaaa | 720 |
| attgaccgga | tgatggaggc | cttcgcccag | cgatactgcc | tgtgcaacct | tgggggtttc | 780 |
| cagtccacag | acacgtgcta | tgtgctgtcc | ttcgccgtca | tcatgctcaa | caccagtctc | 840 |
| cacaatccca | atgtccggga | caagccgggc | ctggagcgct | ttgtggccat | gaaccggggc | 900 |
| atcaacgagg | gcggggacct | gcctgaggag | ctgctcagga | acctgtacga | cagcatccga | 960 |
| aatgagccct | tcaagattcc | tgaggatgac | gggaatgacc | tgaccacac | cttcttcaac | 1020 |
| ccggaccggg | agggctggct | cctgaagctg | ggagggggccc | gggtgaagac | gtggaagcgg | 1080 |
| cgctggttta | tcctcacaga | caactgcctc | tactactttg | agtacaccac | ggacaaggag | 1140 |
| ccccgaggaa | tcacccccct | ggagaatctg | agcatccgag | agggtggacga | cccccgaaa | 1200 |
| ccgaactgct | ttgaacttta | catccccaac | aacaaggggc | aggaggagaa | ggacgagtgg | 1260 |
| atcaagtcca | tccaggcggc | tgtgagtgtg | gaccccttct | atgagatgct | ggcagcgaga | 1320 |
| aagaagcgga | tttcagtcaa | gaagaagcag | gagcagccct | gacccctctg | ccccaaactc | 1380 |
| attatttatt | acggagctgc | ccgcctggg | tggcgggacc | cctgggcctt | ggggctgtgg | 1440 |
| atcctggttc | cctgttttga | aaattcacca | cctctagctc | ctcactgttc | tttgtaatta | 1500 |
| acacgctggt | ggtaatctta | ttaattattt | aaccacttgg | aaaaaaaaa | a | 1551 |

<210> 56

<211> 4254

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510292CB1

<400> 56

| | | | | | | |
|-------------|------------|------------|------------|------------|-------------|-----|
| cgccgcgggag | cgagggttgc | tggagagagc | gcctggggcg | agaagggtta | acggggccacc | 60 |
| gggggctcgc | agagcaggag | ggtgctctcg | gacggtgtgt | ccccactgc | actcctgaac | 120 |
| ttggaggaca | gggtcgccgc | gagggacgca | gagagcacc | tccacgccc | gatgcctgcg | 180 |
| tagtttttgt | gaccagtccg | ctcctgcctc | ccccggggc | agtagagggg | gagcgatgga | 240 |
| gaactggact | ggcaggccct | ggctgtatct | gctgtgctt | ctgtccctcc | ctcagctctg | 300 |
| cttggatcag | gaggtgttgt | ccggacactc | tcttcagaca | cctacagagg | agggccaggg | 360 |
| ccccgaagg | gtctggggac | cttgggtcca | gtgggcctct | tgctcccagc | cctgcggggg | 420 |
| gggggtgcag | cgcaggagcc | ggacatgtca | gtccctaca | gtgcagctcc | acccgagtct | 480 |
| ggccctccct | ccccggcccc | caagacatcc | agaagccctc | ctccccggg | gccagggtcc | 540 |

| | | | | | | |
|-------------|-------------|-------------|-------------|------------|------------|------|
| cagacccacg | acttctccag | aaaccctccc | cttgtacagg | acacagtctc | ggggaagggg | 600 |
| tggcccactt | cgaggtcccc | cttcccacct | agggagagag | gagacccagg | agattcgagc | 660 |
| ggccaggagg | tcccggcttc | gagaccccat | caagccagga | atgttcgggt | atgggagagt | 720 |
| gcccttttgc | ttgccactgc | accggaaccg | caggcaccct | cggagcccac | ccagatctga | 780 |
| gctgtccctg | atctcttcta | gaggggaaga | gcctattccg | tcccctactc | caagagcaga | 840 |
| gccattctcc | gcaaacggca | gccccaaaac | tgagctccct | cccacagaac | tgtctgtcca | 900 |
| caccccatcc | ccccaaagcag | aacctctaag | ccctgaaact | gctcagacag | aggtggcccc | 960 |
| cagaaccagg | cctgcccccc | tacggcatca | ccccagagcc | caggcctctg | gcacagagcc | 1020 |
| cccctcacc | acgcactcct | taggagaagg | tggtctcttc | cgtgcatccc | ctcagccacg | 1080 |
| aaggccaagt | tcccagggtt | gggccagtc | ccaggtagca | gggagacgcc | ctgatccttt | 1140 |
| tccttcgggtc | cctcggggcc | gaggccagca | gggccaaggg | ccttggggaa | cgggggggac | 1200 |
| tcctcacggg | ccccgcctgg | agcctgaccc | tcagcaccgc | ggcgcctggc | tgcccctgct | 1260 |
| gagcaacggc | ccccatgcca | gctccctctg | gagcctcttt | gctcccagta | gccctattcc | 1320 |
| aagatgttct | ggggagagtg | aacagctaag | agcctgcagc | caagcgcctc | gccccctga | 1380 |
| gcagccagac | ccccgggccc | tgcatgccc | agcctttaac | tcccaggaat | tcattggcca | 1440 |
| gctgtatcag | tgggagccct | tcactgaagt | ccagggtctc | cagcgtctgt | aactgaactg | 1500 |
| ccggccccgt | ggcttccgct | tctatgtccg | tcacactgaa | aaggtccagg | atgggaccct | 1560 |
| gtgtcagcct | ggagccccctg | acatctgtgt | ggctggacgc | tgtctgagcc | ccggctgtga | 1620 |
| tgggatcctt | ggctctggca | ggcgtcctga | tggtctgtgga | gtctgtgggg | gtgatgatcc | 1680 |
| tacctgtcgc | cttgtttcgg | ggaacctcac | tgaccagagg | ggccccctgg | gctatcagaa | 1740 |
| gatcttgttg | attccagcgg | gagccttgcc | gctccagatt | gcccagctcc | ggcctagctc | 1800 |
| caactacctg | gcacttctgt | gccctggggg | ccgggtccatc | atcaatggga | actgggctgt | 1860 |
| ggatccccct | gggtcctaca | gggcccggcg | gaccgtcttt | cgatataacc | gtcctcccag | 1920 |
| ggaggagggc | aaaggggaga | gtctgtcggc | tgaaggcccc | accaccacgc | ctgtggatgt | 1980 |
| ctatatgate | tttcaggagg | aaaaccacag | cgttttttat | cagtatgtca | tctcttcacc | 2040 |
| tcctccaate | cttgagaacc | ccacccacga | gccccctgtc | ccccagcttc | agccggagat | 2100 |
| tcaggggtg | gagcccccc | ttgctccggc | accccgcaca | gcccggaccc | caggcacctc | 2160 |
| ccagcgtcag | gtgcggatcc | cccagatgcc | cgccccgcgc | catcccagga | cacccctggg | 2220 |
| gtctccagct | gcgtactgga | aacgagtggg | acactctgca | tgctcagcgt | cctgcgggaa | 2280 |
| aggtgtcttg | cgccccatth | tcctctgcat | ctcccgtgag | tcgggagagg | aactggatga | 2340 |
| acgcagctgt | gcccgcgggtg | ccaggccccc | agcctccccct | gaaccctgcc | acggcacccc | 2400 |
| atgcccccca | tactgggagg | ctggcgagtg | gacatcctgc | agccgctcct | gtggcccccg | 2460 |
| caccagcac | cgccagctgc | agtgcgggca | ggaatttggg | gggggtggct | cctcggtgcc | 2520 |
| cccggagcgc | tgtggacatc | tccccgggcc | caacatcaacc | cagtcttgcc | agctgcgcct | 2580 |
| ctgtggccat | tgggaagtgt | gctctccttg | gagccaggtc | tgggaagccc | agctccctgg | 2640 |
| attccccctg | ccccctcagt | gctccgtgcg | gtgcggccgg | ggccagagaa | gccggcaggt | 2700 |
| tcgctgtgtt | gggaacaacg | gtgatgaagt | gagcgagcag | gagtgtgcgt | caggcccccc | 2760 |
| acagcccccc | agcagagagg | cctgtgacat | ggggccctgt | actactgcct | ggttccacag | 2820 |
| cgactggagc | tcgaagtgtc | cagccgagtg | tgggacggga | atccagcggc | gctctgtggt | 2880 |
| ctgccttggg | agtgccggag | ccctcgggcc | aggccagggg | gaagcaggag | caggaactgc | 2940 |
| gcagagctgt | ccaacaggaa | gccggccccc | tgacatgcgc | gcctgcagcc | tggggccctg | 3000 |
| tgagagaact | tggcgctggt | acacagggcc | ctgggggtgag | tgctcctccg | aatgtggctc | 3060 |
| tggcacacag | cgtagagaca | tcattctgtgt | atccaaactg | gggacggagt | tcaacgtgac | 3120 |
| ttctccgagc | aactgttctc | acctccccag | gccccctgcc | ctgcagccct | gtcaagggca | 3180 |
| ggcctgccag | gaccgatggt | tttccacgcc | ctggagccca | tgctctcgct | cctgccagg | 3240 |
| gggaacgcag | acacgggagg | tccagtgcct | gagcaccac | cagaccctca | gcaccgatg | 3300 |
| ccctcctcaa | ctgcggccct | ccaggaagcg | ccctgtaac | agccaaccct | gcagccagcg | 3360 |
| ccctgatgat | caatgcaagg | acagctctcc | acattgcccc | ctgggtgtac | aggccccgct | 3420 |
| ctgcgtctac | ccctactaca | cagccacctg | ttgccgctct | tgcgcacatg | tcctggagcg | 3480 |
| gtctccccag | gatccctcct | gaaaggggtc | cggggacact | tcacggtttt | ctgtgccacc | 3540 |
| atcggtcacc | cattgatcgg | cccactctga | accccctggc | tctccagcct | gtcccagtct | 3600 |
| cagcagggat | gtcctccagg | tgacagaggg | tggcaaggtg | actgacacaa | agtgaatttc | 3660 |
| agggctgtgg | tcaggcccat | gtggtggtgt | gatgggtgtg | tgacatatg | cctcaggtgt | 3720 |
| gcttttggga | ctgcatggat | atgtgtgtgc | tcaaacgtgt | atcacttttc | aaaaagaggt | 3780 |
| tacacagact | gagaaggaca | agacctgttt | ccttgagact | ttcctaggtg | gaaaggaaag | 3840 |
| caagtctgca | gttccttgct | aatctgagct | acttagagtg | tggctctccc | accaactcca | 3900 |
| gttttgtgcc | ctaagcctca | tttctcatgt | tcagacctca | catcttctaa | gccgcctgt | 3960 |
| gtctctgacc | cttctcatt | tgccctagt | ctctgcccct | gcctccctaa | ttagctagg | 4020 |
| ctggggctcag | ccactgcca | tcctgcctta | ctcaggaagg | caggaggaaa | gagactgcct | 4080 |
| ctccagagca | aggcccagct | gggcagaggg | tgaanaagag | aaatgtgagc | atccgctccc | 4140 |
| ccaccacccc | gcccagcccc | tagcccaact | cctgcctcc | tgaatgggtt | cccacccaga | 4200 |
| actaatattat | tttttattaa | agatggtcat | gacaaatgaa | aaaaaaaaaa | aaaa | 4254 |

<210> 57
 <211> 1509
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7504669CB1

<400> 57
 cggctcgagg cccgagagct ccggggggccg ctgcagccgc ccaagcgccc gccatgcgcg 60
 ctgcccgcgc cgcgcgcgtg ctccagctgc tgctcctgct gggggcgtgg ctggaggctg 120
 cgggcgttgc ggagtcgcgc ctgcccgcgc tggtccttgc catcctggcc cgcaatgccg 180
 aacactcgct gccccactac ctgggcgcctc tggagcggct ggactacccc cgggccagga 240
 tggccctctg gtgtgccacg gaccacaatg tggacaacac cacagagatg ctgcaggagt 300
 ggctggcggc tgtgggcgat gactatgctg ctgtgggtctg gaggcctgag ggcgagccca 360
 ggttctaccc agatgaagag ggtcccaagc actggaccaaa agaaaggcac cagtttctga 420
 tggagctgaa gcaggaagcc ctacaccttg ccaggaactg gggggccgac tatatcctgt 480
 ttgcagacac agacaacatt ctgaccaaca atcagactct gcggcttctc atggggcagg 540
 ggcttccagt ggtggcccca atgctggact ccagaccta ctactccaac ttctgggtgtg 600
 ggatcacccc ccagggtctac taccgccgca cagccgagta cttccccacc aagaaccgcc 660
 agcgcggggg ctgcttccgt gtcccatggg tccactccac cttccttgca tccctgcggg 720
 ctgaagggggc agaccagctt gctttctacc cgccacatcc caactacact tggcctttcg 780
 acgacatcat cgtcttcgcc tatgcctgcc aggtctgctg ggtctccgtc cacgtgtgca 840
 atgagcaccg ttatgggtac atgaatgtgc cggtgaaatc ccaccagggg ctggaagacg 900
 agaggggtcaa cttcatccac ctgatcttag aagcactagt ggacggcccc cgcattgcagg 960
 cctcagctca tgtgactcgg cctcctaaga ggcccagcaa gatagggttt gacagggtct 1020
 ttgtcatcag cctggctcgc aggcctgacc gtcgggaacg catgctcgcc tcgctctggg 1080
 agatggagat ctctgggagg gtggtggacg ctgtggatgg ctggatgctc aacagcagtg 1140
 ccatcaggaa cctggcgcta gacctgctcc cgggctacca ggacccttac tcgggccgca 1200
 ctctgaccaa gggcgagggt ggctgcttcc tcagccatta ctccatctgg gaagagcgag 1260
 cagtacaagg cacacttctg gccacgggac ctggtggcct tctccgcccc gccctgctc 1320
 gctgccccta cccactatgc cggggacgcc gagtggctca gtgacacgga gacatcctct 1380
 ccatgggatg atgacagcgg ccgcctcacc agctggagcg gctcccaaaa gaccctgcgc 1440
 agccccgcgc tggacctgac tggcagcagc gggcacagcc tccaacccca gccccgagat 1500
 gagctctag 1509

<210> 58
 <211> 2439
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7509266CB1

<400> 58
 ggtagcctct gttttcattt cagtcttaat gaaaactttc taacttatat ctcaagtttc 60
 ttttcaaagc agtgtaagta gtatttaaaa tgttatactt caagaaagaa agactttaac 120
 gatattcagc gttgggtctg taacgctgaa ggtaattcat tttttaatcg gtctgcacag 180
 caagaactga aacgaatggg gattgaactg ctttgccctg tctttctatt tctaggaagg 240
 aatgatcacg tacaagggtg ctgtgccctg ggagggtgcag aaacctgtga agactgcctg 300
 cttattggac ctcagtgtgc ctggtgtgct caggagaatt ttactcatcc atctggagtt 360
 ggcgaagggt ggtgcgcaga ctctgcaggg gcatgtccgc cagactgagg actaccgggt 420
 ggatttgtat tacctcatgg acctctccgc ctccatggat gacgacctca acacaataaa 480
 ggagctgggc tcccggcttt ccaaagagat gtctaaatta accagcaact ttagactggg 540
 cttcggatct tttgtggaaa aacctgtatc ccctttttgtg aaaacaacac cagaagaaat 600
 tgccaaccct tgcagtagta ttccatactt ctgtttacct acatttggat tcaagcacat 660
 tttgccattg acaaatgatg ctgaaagatt caatgaaatt gtgaagaatc agaaaatttc 720
 tgctaatatt gacacacccg aagggtggatt tgatgcaatt atgcaagctg ctgtgtgtaa 780
 ggaaaaaatt ggctggcgga atgattccct ccacctcctg gtctttgtga gtgatgctga 840
 ttctcatttt ggaatggaca gcaaactagc aggcacgtgc attcctaatt acggggtctg 900
 tcacttggac agcaagaatg aatactccat gtcaactgtc ttggaatatc caacaattgg 960

```

acaactcatt gataaactgg tacaaaaacaa cgtgttattg atcttcgctg taacccaaga 1020
acaagttcat ttatatgaga attacgcaaa acttattcct ggagctacag taggtctact 1080
tcagaaggac tccggaacaa ttctccagct gatcatctca gcttatgaag aactgcggtc 1140
tgagggtgaa ctggaagtat taggagacac tgaaggactc aacttgctcat ttacagccat 1200
ctgtaacaac ggtaccctct tccaacacca aaagaaatgc tctcacatga aagtgggaga 1260
cacagcttcc ttcagcgtga ctgtgaatat ccacactgc gagagaagaa gcaggcacat 1320
tatcataaag cctgtggggc tgggggatgc cctggaatta cttgtcagcc cagaatgcaa 1380
ctgcgactgt cagaaagaag tggaaagtga cagctccaaa tgtcaccacg ggaacggctc 1440
tttccagtgt ggggtgtgtg cctgccacct ggccacatgg ggcctcgctg tgagtgtggc 1500
gaggacatgc tgagcacaga ttcttgcaag gaggccccag atcatccctc ctgcagcgga 1560
aggggtgact gctactgttg gcagtgtatc tgccacttgt ctccctatgg aaacatttat 1620
gggccttatt gccagtgtga caatttctcc tgcgtgagac acaaagggct gctctgcgga 1680
ggtaacggcg actgtgactg tgggtgaatgt gtgtgcagga gcggctggac tggcgagtac 1740
tgcaactgca ccaccagcac ggactcctgc gtctctgaag atggagtgtc ctgcagcggg 1800
cgcggggact gtgtttgtgg caagtgtgtt tgcacaaacc ctggagcctc aggaccaacc 1860
tgtgaacgat gtcctacctg tggtgacccc tgtaactcta aacggagctg cattgagtgc 1920
cacctgtcag cagctggcca agcccgagaa gaatgtgtgg acaagtgcaa actagctggg 1980
gcgaccatca gtgaagaaga agatttctca aaggatgggt ctgtttcctg ctctctgcaa 2040
ggagaaaatg aatgtcttat tacattccta ataactacag ataacgaggg gaaaaccatc 2100
attcacagca tcaatgaaaa agattgtccg aagcctccaa acattcccat gatcatgtta 2160
ggggtttccc tggtatttct tctcatcggg gttgtcctac tgtgcatctg gaagctactg 2220
gtgtcatctt atgatcgtaa agaagttgccc aaatttgaag cagaacgatc aaaagccaag 2280
tggcaaacgg gaaccaatcc actctacaga ggatccacaa gtacttttaa aaatgtaact 2340
tataaacaca gggaaaaaca aaaggtagac ctttccacag attgctagaa ctactttatg 2400
catgaaaaaa gtctgtttca ctgatatgaa atgttaatg 2439

```

<210> 59

<211> 2440

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509288CB1

<400> 59

```

ggtagcctct gttttcattt cagtcttaat gaaaactttc taacttatat ctcaagtttc 60
ttttcaaagc agtgtaagta gtatttaaaa tgttatactt caagaaagaa agactttaac 120
gatattcagc gttggtcttg taacgctgaa ggtaattcat tttttaatcg gtctgcacag 180
caagaactga aacgaatggg gattgaactg ctttgccctg aaacctgtga agactgcctg 240
aatgatcacg tacaaggtgg ctgtgccctg ggaggtgcag aaacctgtga agactgcctg 300
cttattggac ctcagtgtgc ctggtgtgct caggagaatt ttactcatcc atctggagtt 360
ggcgaaaggt ggtgctcaga ctctgcaggt gcatgtccgc cagactgagg actaccggg 420
ggatttgat tacctcatgg acctctccgc ctccatggat gacgacctca acacaataaa 480
ggagctgggc tcccggcttt ccaaagagat gtctaaatta accagcaact ttagactggg 540
cttcggatct tttgtgga aaacctgtat ccgttttgtg aaaacaacac cagaagaaat 600
tgccaaccct tgcagtagta ttccatactt ctgtttacct acatttggat tcaagcacat 660
tttgccattg acaaatgatg ctgaaagatt caatgaaatt gtgaagaatc agaaaatttc 720
tgctaataat gacacaccg aaggtggatt tgatgcaatt atgcaagctg ctgtgtgtaa 780
ggaaaaaatt ggctggcgga atgattccct ccacctcctg gtctttgtga gtgatgctga 840
ttctcatttt ggaatggaca gcaaactagc aggcatactg attcctaatt acgggctctg 900
tcacttggac agcaagaatg aatactccat gtcaactgtc ttggaatatc caacaattgg 960
acaactcatt gataaactgg tacaaaacaa cgtgttattg atcttcgctg taacccaaga 1020
acaagttcat ttatatgaga attacgcaaa acttattcct ggagctacag taggtctact 1080
tcagaaggac tccggaacaa ttctccagct gatcatctca gcttatgaag aactgcggtc 1140
tgagggtgaa ctggaagtat taggagacac tgaaggactc aacttgctcat ttacagccat 1200
ctgtaacaac ggtaccctct tccaacacca aaagaaatgc tctcacatga aagtgggaga 1260
cacagcttcc ttcagcgtga ctgtgaatat ccacactgc gagagaagaa gcaggcacat 1320
tatcataaag cctgtggggc tgggggatgc cctggaatta cttgtcagcc cagaatgcaa 1380
ctgcgactgt cagaaagaag tggaaagtga cagctccaaa tgtcaccacg ggaacggctc 1440
tttccagtgt ggggtgtgtg cctgccaccc tggccacatg gggcctcgct gtgagtgtgg 1500
cgaggacatg ctgagcacag attcctgcaa ggaggcccca gatcatccct cctgcagcgg 1560
aaggggtgac tgctactgtg ggcagtgtat ctgccacttg tctccctatg gaaacattta 1620

```

```

tgggccttat  tgccagtgtg  acaattttctc  ctgcgtgaga  cacaaagggc  tgctctgcgg  1680
aggtaacggc  gactgtgact  gtggtgaatg  tgtgtgcagg  agcggctgga  ctggcgagta  1740
ctgcaactgc  accaccagca  cggactcctg  cgtctctgaa  gatggagtgc  tctgcagcgg  1800
gcgcggggac  tgtgtttgtg  gcaagtgtgt  ttgcacaaac  cctggagcct  caggaccaac  1860
ctgtgaacga  tgtcctacct  gtggtgaccc  ctgtaactct  aaacggagct  gcattgagtg  1920
ccacctgtca  gcagctggcc  aagcccgaga  agaatgtgtg  gacaagtgca  aactagctgg  1980
tgcgaccatc  agtgaagaag  aagattttctc  aaaggatggt  tctgtttcct  gctctctgca  2040
aggagaaaaa  gaatgtctta  ttacattcct  aataactaca  gataatgagg  ggaaaacct  2100
cattcacagc  atcaatgaaa  aagattgtcc  gaagcctcca  aacattccca  tgatcatggt  2160
aggggtttcc  ctggttatcc  ttctcatcgg  ggttgctcta  ctgtgcatct  ggaagctact  2220
ggtgtcattt  catgatcgta  aagaagttgc  caaatttgaa  gcagaacgat  caaaagccaa  2280
gtggcaaacg  ggaaccaatc  cactctacag  aggatccaca  agtactttta  aaaatgtaac  2340
ttataaacac  agggaaaaac  aaaaggtaga  cttttccaca  gattgctaga  actactttat  2400
gcatgaaaaa  agtctgtttc  actgatatga  aatgttaatg  2440

```

<210> 60

<211> 2580

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510212CB1

<400> 60

```

ggcgccctcc  tcccggggcgg  ataattgaac  ggcgcgggccc  tggccagcgt  tggctgcccga  60
ggctcggccg  gagecgtggag  cccgcgcgcg  tgccccagga  ccgcgcccgc  gcctttgtcc  120
gccgcgcgcc  accgcccgtc  gccgcgcgcg  catggagcgc  gccgcgcgcg  cgcgccgggt  180
cccgctttcc  ctgctgctgc  tcggcggcct  tgcgctgctg  gcggccggag  tggacgcgga  240
tgtcctcctg  gaggcctgct  gtgcggacgg  acaccggatg  gccactcatc  agaaggactg  300
ctcgctgcca  tatgctacgg  aatccaaaga  atgcaggatg  gtgcaggagc  agtgctgcca  360
cagccagctg  gaggagctgc  actgtgccac  gggcatcagc  ctggccaacg  agcaggaccg  420
ctgtgccacg  ccccacggtg  acaacgccag  cctggaggcc  acatttgtga  agaggtgctg  480
ccattgctgt  ctgctgggga  gggcgggcca  ggcccagggc  cagagctgcg  agtacagcct  540
catggttggc  taccagtggt  gacaggtctt  ccgggcatgc  tgtgtcaaga  gccaggagac  600
cggagatttg  gatgtcgggg  gcctccaaga  aacggataag  atcattgagg  ttgaggagga  660
acaagaggac  ccatactga  atgaccgtg  ccgaggaggc  gggccctgca  agcagcagtg  720
ccgagacacg  ggtgacgagg  tggctctgct  ctgcttcgtg  ggctaccagc  tgctgtctga  780
tgggtgtctc  tgtgaagatg  tcaatgaatg  catcacgggc  agccacagct  gccggcttgg  840
agaatcctgc  atcaacacag  tgggctcttt  ccgctgccag  cgggacagca  gctgcgggac  900
tggctatgag  ctcacagagg  acaatagctg  caaagatatt  gacgagtgtg  agagtgggtat  960
tcataactgc  ctccccgatt  ttatctgtca  gaatactctg  ggatccttcc  gctgccgacc  1020
caagctacag  tgcaagagtg  gctttataca  agatgctcta  ggcaactgta  ttgatataca  1080
tgagtgtttg  agtatcagtg  ccccgtgcc  tattgggcat  acatgcatca  acacagaggg  1140
ctcctacacg  tgccagaaga  acgtgcccaa  ctgtggccgt  ggctaccatc  tcaacgagga  1200
gggaacgcgc  tgtgttgatg  tggacgagtg  cgcgccacct  gctgagccct  gtgggaagg  1260
acatcgctgc  gtgaactctc  ccggcagttt  ccgctgcgaa  tgcaagacgg  gttactat  1320
tgacggcatc  agcaggatgt  gtgtcgatgt  caacgagtgc  cagcgctacc  ccgggcgcct  1380
gtgtggccac  aagtgcgaga  acacgctggg  ctccctacct  tgcagctgtt  ccgtgggctt  1440
ccggctctct  gtggatggca  ggtcatgtga  agagagccac  aagtgcgaga  acacgctggg  1500
ctcctacctc  tgcagctgtt  ccgtgggctt  ccggctctct  gtggatggca  ggtcatgtga  1560
agacatcaat  gagtgcagca  gcagccctg  tagccaggag  tgtgccaacg  tctaccgctc  1620
ctaccagtgt  tactgccggc  gaggetacca  gctcagcgat  gtggatggag  tcacctgtga  1680
agacatcgac  gagtgcgccc  tgcccacccg  gggccacatc  tgctcctacc  gctgcatcaa  1740
catccctgga  agcttccagt  gcagctgccc  ctctgtctgg  tacaggctgg  cccccaatgg  1800
ccgcaactgc  caagacattg  atgagtgtgt  gactggcatc  cacaactgct  ccatcaacga  1860
gacctgcttc  aacatccagg  gcggcttccg  ctgcttggcc  ttcgagtggc  ctgagaacta  1920
ccgccgctcc  gcagccacat  gatcgtaggg  aactctgcat  gaggccatcg  gtgcaggctg  1980
gagaagagaa  ggcaagttgg  caggagtgga  gaccacaggc  atttgagcca  ctccctcatg  2040
taacttaact  tgtgccttca  ggacctgtc  aagcccgatc  acgtatatac  cacttccatt  2100
tgatgatgga  atgctgctgt  tcatgaccaa  ctttatggct  agatgggtca  gaaagcacc  2160
agttcatgat  aggcagttca  ggtcatatgg  tgacttgatg  acccagagtc  aaacattcag  2220
ttccaccaa  agcccagtaa  caggccaaga  gctgtctctc  aaaagaagag  tagttatctg  2280

```



```

cagaagatgg cagggccttg ctccgaaagc ctagagaccg ccactgtgat tcacctatgg 2340
gggcctgcc aagctgcagc cagcatcctt atctgccact gacacctcaa gcaacattgg 2400
atctgctggg tcatatggcc caagtggcag agcaacttgc acaacagcct ggacctgtca 2460
tagagctttc tcctgttctg gacccactc aaaactggca gcctttcagg tcactcaata 2520
aatgtgctgg agtaacacac tcaaacgagg aatgtgttgc ctccaaaatc caataggccc 2580

```

```

<210> 61
<211> 8029
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7510504CB1

```

```

<400> 61
tttcccttct cgtctctgtg ctgctctgcc tttctttttc gtcctttttc cccttttttc 60
tcgcaccgtt aacttttctc tgtcctcctg gcttggttatt cagcatgtct gattaacagg 120
agcctgattg gctggctgcc tgctccgaga gagattccat tcagcttacc cccaccccat 180
ccaccacccc acccctcggc caggaaatgt gagaggggct gatggaagct gataggcagg 240
actggagtgt tagcaccagt actggatgtg acagcaggca gaggagcact tagcagctta 300
ttcagtgtcc gattctgatt ccggcaagga tccaagcatg gaatgctgcc gtcggggcaac 360
tcctggcaca ctgctcctct tcttggtttt cctgctcctg agttccagga ccgcacgctc 420
cgaggaggac cgggacggcc tatgggatgc ctggggccca tggagtgaat gctcacgcac 480
ctgcggggga ggggcctcct actctctgag gcgctgcctg agcagcaaga gctgtgaagg 540
aagaaatata cgatacagaa catgcagtaa tgtggactgc ccaccagaag cagggtgattt 600
ccgagctcag caatgctcag ctcataatga tgtcaagcac catggccagt tttatgaatg 660
gcttctgtgt tctaattgacc ctgacaaccc atgttcactc aagtgccaaag ccaagggaac 720
aaccttggtt gttgaactag cacctaaggt cttagatggc acgcgttgct atacagaatc 780
tttggtatat tgcacagtg gtttatgcca aattgttggc tgcgatcacc agctgggaag 840
caccgtcaag gaagataact gtggggtctg caacggagat ggggccacct gccggctggc 900
ccgagggcag tataaatccc agctctccgc aaccaaactc gatgatactg tggttgcaat 960
tccttatgga agtagacata ttgccttgt cttaaaaggc cctgatcact tatactctga 1020
aaccaaaacc ctccagggga cttaaaggctg aaacagctct agctccacag gaactttcct 1080
tgtggacaat tctagtgtgg acttcagaa atttcagac aaagagatac tgagaatggc 1140
tggaacctc acagcagatt tcattgtcaa gattcgtaac tcgggctccg ctgacagtac 1200
agtccagttc atcttctatc aacccatcat ccaccgatgg agggagacgg atttctttcc 1260
ttgctcagca acctgtggag gaggttatca gctgacatcg gctgagtgtc acgatctgag 1320
gagcaaccgt gtggttgctg accaatactg tcaactattac ccagagaaca tcaaacccaa 1380
acccaagctt caggagtga acttggatcc ttgtccagcc agtgacggat acaagcagat 1440
catgccttat gacctctacc atccccttcc tcggtgggag gccaccccat ggaccgctg 1500
ctcctcctcg tgtggggggg gcatccagag ccgggcagtt tcctgtgtgg aggaggacat 1560
ccagggggcat gtcacttcag tgggaagagt gaaatgcatg tacacccta agatgcccat 1620
cgcgcagccc tgcaacattt ttgactgccc taaatggctg gcacaggagt ggtctccgtg 1680
cacagtgaca tgtggccagg gcctcagata ccgtgtggtc ctctgcatcg accatcgagg 1740
aatgcacaca ggaggctgta gcccacaaac aaagccccc ataaaagagg aatgcacgt 1800
accactccc tgctataaac ccaaagagaa acttccagtc gaggccaagt tgccatgggt 1860
caaacaagct caagagctag aagaaggagc tgcgtgtgtc agggagccct cgttcatccc 1920
agaggcctgg tcggcctgca cagtcacctg tgggtgtggg acccaggtgc gaatagtcag 1980
gtgccaggtg ctctgtctt tctctcagtc cgtggctgac ctgcctattg acgagtgtga 2040
agggcccaag ccagcatccc agcgtgctc tcatgcaggc ccatgcagcg gggaaattcc 2100
tgagttcaac ccagacgaga cagatgggct ctttgggtgg ctgcaggatt tcgacgagct 2160
gtatgactgg gagtatgagg ggttcaccaa gtgctccgag tcctgtggag gaggtgtcca 2220
ggaggctgtg gtgagctgct tgaacaaaca gactcgggag cctgctgagg agaacctgtg 2280
cgtgaccagc cgccggcccc cacagctcct gaagtcctgc aatttggatc cctgccagc 2340
aaggtgggaa attggcaagt ggagtccatg tagtctcaca tgtggggtcg gcctacagac 2400
cagagacgtc ttctgcagcc acctgctttc cagagagatg aatgaaacag tcactcctggc 2460
tgatgagctg tgtcggccagc ccaagcccag cacggtgcaa gcttctaacc gctttaattg 2520
ccccagacc tggtaacctg cacagtggca gccgtgttcc agaacgtgtg gcgggggtgt 2580
tcagaaacgt gaggttcttt gcaagcagcg catggctgat ggcagcttcc tggagcttcc 2640
tgagaccttc tgttcagctt caaaacctgc ctgccagcaa gcatgcaaga aagatgactg 2700
tcccagcgag tggcttctct cagactggac agagtgttcc acaagctgcg gggaaggcac 2760
ccagactcga agcgccattt gccgaaagat gctgaaaacc ggcctctcaa cggttgtcaa 2820

```

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tccaccctg | tgcccgcccc | tgcctttctc | ttcctccatc | aggccctgta | tgctggcaac | 2880 |
| ctgtgcaagg | cccgggcggc | catccacgaa | gcacagcccc | cacatcgcg | ccgccaggaa | 2940 |
| ggctctacatc | cagactcgca | ggcagaggaa | gctgcacttc | gtggtggggg | gcttcgccta | 3000 |
| cctgctcccc | aagacggcgg | tgggtgctgcg | ctgcccggcg | cgaggggtcc | gcaagcccc | 3060 |
| catcacctgg | gagaaggacg | gccagcacct | catcagctcg | acgcacgtca | cggtggcccc | 3120 |
| cttcggctat | ctcaagatcc | accgcctcaa | gccctcggat | gcaggcgtct | acacctgctc | 3180 |
| agcgggcccc | gcccgggagc | actttgtgat | taagctcatc | ggaggcaacc | gcaagctcgt | 3240 |
| ggccccggccc | ttgagccccg | gaagtgagga | agaggtgctt | gcggggagga | agggcgggccc | 3300 |
| gaaggaggcc | ctgcagaccc | acaaacacca | gaacgggata | ttctccaacg | gcagcaaggc | 3360 |
| ggagaagcgg | ggcctggcgg | ccaacccggg | gagccgctac | gacgacctcg | tctcccggct | 3420 |
| gctggagcag | ggcggctggc | ccggagagct | gctggcctcg | tgggagggcg | aggactccgc | 3480 |
| ggaaaggaac | acgacctcgg | aggaggaccc | gggtgcagag | caagtgtctc | tgcacctgcc | 3540 |
| cttcaccatg | gtgaccgagc | agcggcgcc | ggacgacatc | ctgggggaacc | tctcccagca | 3600 |
| gcccagaggag | ctgcgcgacc | tctacagcaa | gcacctgggtg | gcccagctgg | cccaggagat | 3660 |
| cttcgcgacg | cacctggagc | accaggacac | gctcctgaag | ccctcggagc | gcaggaccttc | 3720 |
| cccagtgact | ctctcgcttc | ataaacacgt | gtctggcttc | agcagctccc | tgccggacctc | 3780 |
| ctccaccggg | gacgccgggg | gaggctctcg | aaggccacac | cgcaagccca | ccatcctgcg | 3840 |
| caagatctca | gcggcccagc | agctctcagc | ctcggagggtg | gtcaccaccc | tggggcagac | 3900 |
| ggtggccctg | gccagcggga | cactgagtgt | tcttctgcac | tgtgaggcca | tccggccacc | 3960 |
| aaggcctacc | atcagctggg | ccaggaatgg | agaagaagtt | cagttcagtg | acaggattct | 4020 |
| tctacagcca | gatgatctct | tacagatctt | ggcaccagtg | gaagcagatg | tgggtttctta | 4080 |
| cacttgcaat | gtccaccaatg | ccttgggata | cgactctgtc | tccattgccg | tcacattagc | 4140 |
| aggaaagcca | ctagtgaata | cgtcacgaat | gacagtgatc | aacacggaga | agcctgcagt | 4200 |
| cacagtgcgat | ataggaagca | ccatcaaaac | agtgcaggga | gtgaatgtga | caatcaactg | 4260 |
| ccagggttgca | ggagtgcctg | aagctgaagt | cacttggttc | aggaataaaa | gcaaactggg | 4320 |
| ctccccgcac | catctgcacg | aaggctcctt | gctgctcaca | aacgtgtcct | cctcggatca | 4380 |
| gggcctgtac | tcttgacagg | cggccaatct | tcatggagag | ctgactgaga | gcaccacctc | 4440 |
| tgctgatccta | gatccccccc | aagtctggaa | gacatcaggg | gacatcaggg | ccttgctcgc | 4500 |
| tgccactgga | ccgaaccttc | cttcagtgtc | gacgtctcct | ctgggaacac | agctggctct | 4560 |
| ggatcctggg | aattctgtct | tccttggtcg | ccccatcaaa | ggtcaccctg | tccctaatat | 4620 |
| cacctgggtt | catgggtggt | agccaattgt | cactgccaca | ggactgacgc | atcacatctt | 4680 |
| ggcagctgga | cagatccttc | aagttgcaaa | ccttagcggt | gggtctcaag | gggaattcag | 4740 |
| ctgccttgct | cagaatgagg | caggggtgtc | catgcagaag | gcactcttag | tgatccaagt | 4800 |
| ttactggtgg | tctgtggaca | gactggcaac | ctgctcagcc | tcctgtggta | accggggggt | 4860 |
| tcagcagccc | cgcttgaggt | gcctgctgaa | cagcacggag | gtcaaccctg | cccactgcgc | 4920 |
| agggaagggt | cgccctgcgg | tgcagcccat | cgctgcaac | cggagagact | gcccttctcg | 4980 |
| gtggatggtg | acctcctggt | ctgcctgtac | ccggagctgt | gggggaggtg | tccagacccg | 5040 |
| cagggtgacc | tgtcaaaagc | tgaagccctc | tgggatctcc | accctgtgtg | ccaatgacat | 5100 |
| gtgcaccag | gtcgccaagc | ggcctgtgga | caccagggc | tgtaccagc | agctgtgtgt | 5160 |
| ggagtggggc | ttctccagct | ggggccagtg | caataggcct | tgcatcgggc | ctcacctagc | 5220 |
| tgtgcaacac | agacaagtct | tctgccagac | acgggatggc | atcaccttac | catcagagca | 5280 |
| gtgcagtgtc | cttcagaggc | ctgtgagcac | ccagaactgc | tggtcagagg | cctgcagtgt | 5340 |
| acactggaga | gtcagcctgt | ggaccctgtg | cacagctacc | tgtggcaact | acggcttcca | 5400 |
| gtcccggcgt | gtggagtgtg | tgcattgccg | caccaacaag | gcagtgcctg | agcacctgtg | 5460 |
| ctcctggggg | ccccggcctg | ccaactggca | gcgctgcaac | atcaccccat | gtgaaaacat | 5520 |
| ggagtgcaga | gacaccacca | ggtactgcga | gaagtgaaa | cagctgaaac | tctgccaact | 5580 |
| cagccagttt | aaatctcgct | gctgtggaac | ttgtggcaaa | gcgtgaagat | aggggtgtggg | 5640 |
| gaaaaactct | accctggcca | cacgaaggac | tcacgcaacc | acctcggaca | gaacctaaagc | 5700 |
| tttcttcatt | ttattttatt | atttccccct | ccccactcca | cacacaccct | tccaacctcc | 5760 |
| tccacctcca | ccttcaagca | taaggacgtc | cgcgtgtttt | ctctttcagt | tagctggagg | 5820 |
| acaggatgtt | gggaaaggaa | aggacagatg | tctaaaggag | gttgacagag | aggccaggca | 5880 |
| gacagtgggg | gtcctcttga | agagcttcct | ccctcccaaa | cctgggtctc | aaagacctag | 5940 |
| aaagaggcag | gcacagcccc | tgcggacagc | agggagccag | aagggttgta | gcctattgggt | 6000 |
| gcaaacattg | gacaaattcc | tgtgtctttc | ctagaagcgc | actatcaca | acacaggagt | 6060 |
| gttttgctcc | tttgtctcct | cttccccatc | tatgtccctt | tagtcacagt | taggacaaat | 6120 |
| ggggagggga | caccatgctg | aggcagaaac | tagcccagaa | ctcactcagt | tcttctagt | 6180 |
| ggtgagtgc | gagagagaag | aactcagatc | accagtaggg | agaggtaaaa | aagcaaacaa | 6240 |
| agcaggctct | aaggcacaca | acattgcaga | aaatgaggaa | gggaggggag | ggaagggaca | 6300 |
| gaagcaaaaa | ggagcctgtg | gtgttcccca | gtggggcagg | gtgagcaggg | gttccaggc | 6360 |
| tgcatgagc | tcatggacca | gctctgatcc | catgcatgtg | cgcatgctca | gagccctgct | 6420 |
| gccacaaca | gagcactgcg | ctgcgtggga | gtccccactt | cccaagctat | cagagtcaac | 6480 |
| gtcctgcctg | tgcagctgca | gcaaagccag | tgagaggtgg | gtctcgccat | gcagtgaaggc | 6540 |
| cacctgggca | cctctttatc | taaatccgaa | gtccccctagc | ccgcactaa | ctaactgctg | 6600 |

| | | | | | | |
|-------------|------------|-------------|-------------|-------------|-------------|------|
| ctgtggggcca | gggccatttt | gagcatgaat | ggcccagggt | ttttgccttc | taggaccttt | 6660 |
| gctgctccac | cgaagggcca | gggactatgg | ttaacttatc | aacatcaacc | cattaactag | 6720 |
| tactgtgccc | agagagtatc | tgtcaggctg | tcaggttgta | gcaacctctt | cattccagag | 6780 |
| ctggcccagg | gaccggggtg | ggacaatggg | tttatgcgtg | tccacagtac | acctccctc | 6840 |
| tcccagcctc | caccccgagg | tctgcaggtc | ctccggcatg | tagtatttat | ctagcaaggc | 6900 |
| ggggtggtgg | aggcagcacc | ctggcaaagc | agctcacaca | ctgcagccac | actcatcagc | 6960 |
| tgtggtgagg | cggctggagc | aaagtcaaag | tcatgcagca | aaatgaaaac | tctgggactc | 7020 |
| ttcggcaaaa | tcctcattaa | gccgagcagc | tttggccaag | taattttttg | ctccttccct | 7080 |
| cgcgtggcct | gagtttagga | gcaaggggtg | ccagagtccc | ttaccacacag | ataagcctcc | 7140 |
| cctcatgaaa | tgccactcac | cccgggctac | cattgacatc | agggctgcat | ttccagccag | 7200 |
| cctggaagta | aaatttgaga | ggaagacaat | attaatctgt | gtccccacct | agtgaagctgt | 7260 |
| ggacagggtt | aagttgggtc | tccttcttct | tcaccacaaa | aacaggctct | aagaaatcat | 7320 |
| gttactaaaa | aatcagtgta | aagtctgttt | aaaataaaaa | agaatgtttt | ctatgtctgt | 7380 |
| atatcttttg | tgaatattta | ttaggatttc | ttattaaaaa | agtgcataat | taataattgt | 7440 |
| acattgtcat | ccagaaacaa | aactattggg | gggactttat | taactaactt | cctgcagttg | 7500 |
| tgttcctgta | aactcagtag | tgattattat | atttttcccta | tttttaatat | aacctgggtg | 7560 |
| ttaaactctgg | atccattcac | tgtacaggat | gtgttgtaaa | aactaacatg | ggatgctgag | 7620 |
| gcagtaagag | ggaattcatt | tgtggcataa | tagttatgca | tggaaatgata | aagacagaca | 7680 |
| aattccatac | tactactaat | gtgggttaatt | atttctagtt | cgatagtgat | tgaaaatcag | 7740 |
| tggtcactat | ttacatttcc | taaagagcaa | gcattcctcca | gctccatggt | gggttggagc | 7800 |
| agttggcagt | gggtctcagt | gagctggcag | aacctaggtt | tgggtgggaa | gcagaatgct | 7860 |
| cgttgcatga | aatgaatgta | catttaattgt | ttgttctgtg | aattgcaact | cagcagcacc | 7920 |
| acaagacaat | gaaggctgct | ggctaattgtg | gaaggaggca | ctttctcctc | taaaacacaa | 7980 |
| aactgtattt | gtattttttg | tacagataat | acagcttatt | tattttaat | | 8029 |

<210> 62

<211> 1868

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510587CB1

<400> 62

| | | | | | | |
|-------------|-------------|------------|------------|-------------|-------------|------|
| cttaatgact | agaattcagg | ttccaaggag | aagcccacaa | ggctaagggt | attggatata | 60 |
| acggaaagtg | gaagctatac | ctgacttcca | gagaatgtgg | accggatata | agatcttaat | 120 |
| cttctcttat | cttactacag | aaatctggat | ggagaagcag | tatttatctc | aaagagaagt | 180 |
| ggacctagag | gcttattttc | ctaggaatca | caccgttttg | caaggctactc | gattcaaaaag | 240 |
| agccattttc | caagggcaat | actgtagaaa | ttttggctgt | tgtgaagaca | gagatgatgg | 300 |
| ctgtgtcact | gagttctatg | cggcgaatgc | gttggtgtac | tgtgataaat | tctgtgacag | 360 |
| agaaaattct | gattgctgtc | ctgactacaa | gtcctttttg | cgtgaagaga | aagaatggcc | 420 |
| tcctcacaca | cagccttggt | atccagaagg | ttgcttcaaa | gatgggtcaac | attatgaaga | 480 |
| gggatcagta | attaaagaaa | actgcaactc | ctgcacatgc | tcaggacagc | aatggaaaatg | 540 |
| ttcccagcat | gtatgccttg | ttcgtccaga | attaattgaa | caggtcaata | aaggagacta | 600 |
| tggatggaca | gcacagaatt | acagccaatt | ttggggaatg | actttagaag | atggttttaa | 660 |
| atttcgcctt | ggcacttttg | cacctagtcc | catgctcctg | agcatgaatg | aaatgcagc | 720 |
| ttctttacct | gcaacaactg | atcttccaga | gttttttgtt | gcttcttata | aatggcctgg | 780 |
| atggactcat | ggcccatttg | atcaaaaaaa | ttgtgctgca | tcctgggcat | tttccactgc | 840 |
| aagtgtggct | gctgaccgaa | tagcaattca | gtctaagggt | cgatacacgg | ccaatctatc | 900 |
| ccctcagaat | ttgatctctt | gctgtgccaa | gaaccgtcat | ggatgcaata | gtggaagcat | 960 |
| cgatagggtc | tggttggtacc | tgagaaaacg | tggatggta | tcccacgcac | gctaccact | 1020 |
| tttcaaagac | caaaatgcta | ccaacaatgg | atgtgccatg | gcaagcagg | ctgatgggag | 1080 |
| aggaaaacgg | catgccacga | agccatgtcc | caacaacgta | gaaaaatcta | acaggatcta | 1140 |
| tcaatgttct | cctccataca | gagtctcttc | caacgaaact | gagataatga | aagaaatcat | 1200 |
| gcaaaatgga | ccagttcaag | ccataatgca | agtccgtgaa | gatttcttcc | attataagac | 1260 |
| agggatatac | agacatgtta | ccagcacaaa | taaagaatca | gaaaaatatc | gaaagcttca | 1320 |
| gacacatgca | gtcaaactca | ctggattgct | gccaatctct | ggggaaagtc | atggggagag | 1380 |
| aatggctatt | tcaggattct | tcgaggagta | aatgagtcgc | acattgaaaa | gttgattatc | 1440 |
| gcagcttggg | gccaaactgac | gagttctgat | gaaccataac | atatcattaa | atttccataa | 1500 |
| ggtcatgcct | ttaagtaacc | ccctaaattg | aagtttagca | atatgacatt | cttggtgaca | 1560 |
| gtggaatctt | tgtctcttca | ccgtgttaac | ataatctatc | tattttctta | ttttccctc | 1620 |
| tggctctatgc | ttctgcttcc | ttcatattac | tgagcattaa | caacaccaat | aaaggacagc | 1680 |

```

agagtcacctt aatgtcaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1740
aggggggggcc caaaaaaagg gaaccaaacc cccgggaaaa aaatcggggg ggggacctgg 1800
ggggggaaaaa attttccctta aaggggggggc taaaaaaagg tgggggaaaat gggaaaaagt 1860
tttttggg

```

```

<210> 63
<211> 1531
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7510684CB1

```

```

<400> 63
cttaatgact agaattcagg ttccaaggag aagcccacaa ggctaagggt attggatata 60
acggaaaagt gaagctatac ctgacttcca gagaatgtgg accggatata agatcttaat 120
cttctcttat cttactacag aaatctggat ggagaagcag tatttatctc aaagagaagt 180
ggacctagag gcttattttca ctaggaatca caccgttttg caagggtgct tcaaagatgg 240
tcaacattat gaagagggat cagtaattaa agaaaactgc aactcctgca catgctcagg 300
acagcaatgg aaatgttccc agcatgtatg ccttggtcgt tcagaattaa ttgaacagggt 360
caataaagga gactatggat ggacagcaca gaattacagc caattttggg gaatgacttt 420
agaagatggg tttaaaatttc gccttggcac tttgccacct agtcccatgc tcctgagcat 480
gaatgaaatg acagcttctt tacctgcaac aactgatctt ccagagtttt ttgttgcttc 540
ttataaatgg cctggatgga ctcatggccc attggatcaa aaaaattgtg ctgcatcctg 600
ggcattttcc actgcaagtg tggctgctga ccgaatagca attcagtcta agggctcgata 660
cacggccaat ctatccccctc agaatttgat ctcttgctgt gccaagaacc gtcattggatg 720
caatagtgga agcatcgata gggcttggtg gtacctgaga aaacgtggac tggatatcca 780
cgcatgctac ccacttttca aagacaaaaa tgctaccaac aatggatgtg ccatggcaag 840
caggtctgat gggcgaggaa aacggcatgc cacgaagcca tgtcccaaca acgtagaaaa 900
atctaacagg atctatcaat gttctcctcc atacagagtc tcttccaacg aaactgagat 960
aatgaaagaa atcatgcaaa atggaccagt tcaagccata atgcaagtcc gtgaagattt 1020
cttccattat aagacaggga tatacagaca tgttaccagc acaaataaag aatcagaaaa 1080
atatcgaaag cttcagacac atgcagtcga actcactgga tggggcacac tgagaggagc 1140
acaagggcag aaagaaaaat tttggattgc tgccaattcc tgggggaaagt catggggaga 1200
gaatggctat ttcaggattc ttcgaggagt aaatgagtcc gacattgaaa agttgattat 1260
cgcagcttgg ggccaactga cgagttctga tgaaccataa catatcatta aatttccata 1320
aggatcatgcc tttaagtaac cccctaaatt gaagttagc aatatgacat tcttggtgac 1380
agtggaaatct ttgtctcttc accgtgttaa cataatctat ctattttctt attttcccct 1440
ctggtctatg cttctgcttc cttcatatta ctgagcatta acaacaccaa taaaggacag 1500
cagagtccct aaatgtcttt aaagttaaaa a
1531

```

```

<210> 64
<211> 4220
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 7510697CB1

```

```

<400> 64
gtggccgccc cggagcggag ttgcctggag agagcgccct ggccgcagaag ggttaacggg 60
ccaccggggg ctgcagagc agggagggtg tctcggacgg tgtgtccccc actgcactcc 120
tgaacttggg ggacaggggt gccgcgaggg acgcagagag caccctccac gccagatgc 180
ctgcgtagtt tttgtgacca gtccgctcct gcctccccct ggggcagtag agggggagcg 240
atggagaact ggactggcag gccctggctg tatctgctgc tgccttctgtc cctccctcag 300
ctctgcttgg atcaggagggt gttgtccgga cactctcttc agacacctac agaggagggc 360
cagggccccc aaggtgtctg gggaccttgg gtccagtggg cctcttgctc ccagccctgc 420
ggggtggggg tgcagcgag gagccggaca tgtcagctcc ctacagtgca gctccacccg 480
agtctgcccc tccctccccg gcccccaaga catccagaag cctcctccc ccggggccag 540
ggtcccagac cccagacttc tccagaaacc ctccccttgt acaggacaca gtctcgggga 600
aggggtggcc cacttcgagg tcccgtctcc cacctaggga gagaggagac ccaggagatt 660

```

| | | | | | | |
|------------|-------------|-------------|------------|-------------|-------------|------|
| cgagcggcca | ggaggtcccg | gcttcgagac | cccatcaage | caggaatggt | cggttatggg | 720 |
| agagtgccct | ttgcattgcc | actgcaccgg | aaccgcaggc | accctcggag | cccaccaga | 780 |
| tctgagctgt | ccctgatctc | ttctagaggg | gaagaggcta | ttccgtcccc | tactccaaga | 840 |
| gcagagccat | tctccgcaaa | cggcagcccc | caaactgagc | tccctcccac | agaactgtct | 900 |
| gtccacaccc | catcccccca | agcagaacct | ctaagccctg | aaactgctca | gacagaggtg | 960 |
| gccccagaa | ccaggcctgc | ccccctacgg | catcacccca | gagcccaggc | ctctggcaca | 1020 |
| gagccccct | caccacagca | ctccttagga | gaaggtggct | tcttccgtgc | atccccctag | 1080 |
| ccacgaaggc | caagttccca | gggttggggc | agtccccagg | tagcagggag | acgcctgat | 1140 |
| ccttttctct | cgggtccctcg | gggcccaggc | cagcagggcc | aagggccttg | gggaacgggg | 1200 |
| gggactcctc | acgggccccg | cctggagcct | gacctcagc | acccgggcgc | ctggctgccc | 1260 |
| ctgctgagca | acggccccca | tgccagctcc | ctctggagcc | tctttgctcc | cagtagccct | 1320 |
| attccaagat | gtttctgggga | gagtgaacag | ctaagagcct | gcagccaagc | gccctgcccc | 1380 |
| cctgagcagc | cagacccccg | ggccctgcag | tgccgagcct | ttaactccca | ggaattcatg | 1440 |
| ggccagctgt | atcagtgagg | gcccttcaact | gaagtccagg | gctcccagcg | ctgtgaactg | 1500 |
| aactgccggc | cccggtggctt | ccgcttctat | gtccgtcaca | ctgaaaagg | ccaggatggg | 1560 |
| accctgtgtc | ccctggagc | ccctgacatc | tgtgtggctg | gacgctgtct | gagccccggc | 1620 |
| tgtgatggga | tccttggtct | tgccagggct | cctgatggct | gtggagctct | tgggggtgat | 1680 |
| gattctacct | gtcgcttctg | ttcgggggaa | ctcactgacc | gagggggccc | cctgggctat | 1740 |
| cagaagatct | tgtggattcc | agcgggagcc | ttgcggctcc | agattgcccc | gctccggcct | 1800 |
| agctccaact | acctggcact | tcgtggccct | ggggggccgt | ccatcatcaa | tgggaactgg | 1860 |
| gctgtggatc | cccctgggtc | ctacagggcc | ggcgggaccg | tctttcgata | taaccgtcct | 1920 |
| cccaggagg | aggggcaagg | ggagagtctg | tcggctgaag | gccccaccac | ccagcctgtg | 1980 |
| gatgtctata | tgatctttca | ggaggaaaac | ccaggcgctt | tttatcagta | tgtcatctct | 2040 |
| tcacctctc | caatccttga | gaacccacc | ccagagcccc | ctgtccccc | gcttcagccg | 2100 |
| gagattctga | gggtggagcc | cccacttgct | ccggcaccac | gcccagcccc | gaccccaggc | 2160 |
| accctccagc | gtcaggtgct | gatccccag | atgcccgcct | cgccccatcc | caggacaccc | 2220 |
| ctgggggtct | cagctgcgta | ctggaaacga | gtgggacact | ctgcatgctc | agcgtcctcc | 2280 |
| gggaaagggt | tctggcgccc | cattttctct | tgcatctccc | gtgagtcggg | agaggaactg | 2340 |
| gatgaacgca | gctgtgccgc | gggtgccagg | ccccagcct | cccctgaacc | ctgccacggc | 2400 |
| accccatgcc | ccccatactg | ggaggctggc | gagtggacat | cctgcagccg | ctcctgtggc | 2460 |
| cccggcacc | agcaccgcca | gctgcagtgc | cggcaggaat | ttgggggggg | tggctcctcg | 2520 |
| gtgcccccg | agcgtctgtg | acatctcccc | cggcccaaca | tcaccagtc | ttgccagctg | 2580 |
| cgccctctgt | gccattggga | agttggctct | ccttggagcc | agtgtccctg | gcgggtgcggc | 2640 |
| cggggccaga | gaagccggca | ggttcgctgt | gttgggaaca | atggtgatga | agtgagcgag | 2700 |
| caggagtgtg | cgtcaggccc | cccacagcct | ccagcagag | aggcctgtga | catggggccc | 2760 |
| tgtactactg | cctgggttcca | cagcgactgg | agctccaagt | gctcagccga | gtgtgggacg | 2820 |
| ggaatccagc | ggcgctctgt | ggtctgcctt | gggagtgggg | cagccctcgg | gccaggccag | 2880 |
| ggggaagcag | gagcaggaac | tgggcagagc | tgtccaacag | gaagccggcc | ccctgacatg | 2940 |
| cgcgcctgca | gcctggggcc | ctgtgagaga | acttggcgct | ggtacacagg | gccctggggg | 3000 |
| gagtgtcctc | ccgaatgtgg | ctctggcaca | cagcgtagag | acatcatctg | tgtatccaaa | 3060 |
| ctggggagcg | agttcaacgt | gacttctccg | agcaactgtt | ctcacctccc | caggccccct | 3120 |
| gccctgcagc | cctgtcaagg | gcaggcctgc | caggaccgat | ggttttccac | gccctggagc | 3180 |
| ccatgttctc | gctcctgcca | agggggaacg | cagacacggg | aggtccagtg | cctgagcacc | 3240 |
| aaccagaccc | tcagaccccg | atgccctcct | caactgcggc | cctccaggaa | gcgccccctg | 3300 |
| aacagccaac | cctgcagcca | gcgcctctat | gatcaatgca | aggacagctc | tccacattgc | 3360 |
| cccctggtgg | tacaggcccc | gctctgcgtc | taccctact | acacagccac | ctggtgccgc | 3420 |
| tcttgcgcac | atgtcctgga | gcggctctcc | caggatccct | cctgaaaggg | gtccggggca | 3480 |
| ccttcacggg | tttctgtgcc | accatcggtc | acccattgat | cggccactc | tgaaccccc | 3540 |
| ggctctccag | cctgtcccag | tctcagcagg | gatgtcctcc | aggtgacaga | gggtggcaag | 3600 |
| gtgactgaca | caaagtgact | ttcagggtcg | tggtcaggcc | catgtggtgg | tgtgatgggt | 3660 |
| gtgtgcacat | atgcctcagg | tgtgcttttg | ggactgcatg | gatatgtgtg | tgtcaaacg | 3720 |
| tgtatcactt | ttcaaaaaga | ggttacacag | actgagaagg | acaagacctg | tttcttgtag | 3780 |
| actttcctag | gtggaaagga | aagcaagtct | gcagttcctt | gctaactctga | gctacttaga | 3840 |
| gtgtgggtct | cccaccaact | ccagttttgt | gccctaagcc | tcatttctca | tgttcagacc | 3900 |
| tcacatcttc | taagccgccc | tgtgtctctg | accccttctc | atttgcctag | tatctctgcc | 3960 |
| cctgcctccc | taattagcta | gggctggggg | cagccactgc | caatcctgcc | ttactcagga | 4020 |
| aggcaggagg | aaagagactg | cctctccaga | gcaaggccca | gctgggcaga | gggtgaaaaa | 4080 |
| gagaaatgtg | agcatccgct | ccccaccac | cccggccagc | ccctagcccc | actccctgcc | 4140 |
| tctgaaatg | gttcccacc | agaactaatt | tattttttat | taaagatggg | catgacaaat | 4200 |
| gaaaaaaaaa | aaaaaaaaaa | | | | | 4220 |

<210> 65

<211> 5353

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7761337CB1

<400> 65

```

ccagccaaac ccacctccac catggggggcc atgactcagc tgttggcagg tgtctttctt 60
gcttttccttg ccctcgctac cgaagggtggg gtcctcaaga aagtcatccg gcacaagcga 120
cagagtggggg tgaacgccac cctgccagaa gagaaccagc cagtgggtgt taaccacgtt 180
tacaacatca agctgccagt gggatcccag tgttcggtgg atctggagtc agccagtggg 240
gagaaagacc tggcaccgcc ttcagagccc agcgaaagct ttcaggagca cacagtagat 300
ggggaaaacc agattgtctt cacacatcgc atcaacatcc cccgccgggc ctgtggctgt 360
gccgcagccc ctgatgttaa ggagctgctg agcagactgg aggagctgga gaacctggtg 420
tcttccctga gggagcaatg tactgcagga gcaggctgct gtctccagcc tgccacaggc 480
cgcttggaca ccaggccctt ctgtagcggg cggggcaact tcagcactga aggatgtggc 540
tgtgtctgcg aacctggctg gaaaggcccc aactgctctg agcccgaatg tccaggcaac 600
tgtcaccttc gaggccggtg cattgatggg cagtgcattc gtgacgacgg cttcacgggc 660
gaggactgca gccagctggc ttgccccagc gactgcaatg accagggcaa gtgctgtaat 720
ggagtctgca tctgtttcga aggtacgccc ggggctgact gcagccgtga aatctgcca 780
gtgccctgca gtgaggagca cggcacatgt gtatagtggt tgtgtgtgtg ccacgatggc 840
tttgaggcgg atgactgcaa caagcctctg tgtctcaaca attgctacaa ccgtggacga 900
tgctgtggaga atgagtgcgt gtgtgatgag ggtttcacgg gcgaagactg cagtgaactc 960
atctgcccc aatgactgct cgaccggggc cgctgcattc atggcacctg ctactgcgaa 1020
gaaggcttca caggtgaaga ctgcccggaa cccacctgcc cacatgcctg ccacaccag 1080
ggccgggtgt aggaggggca gtgtgtatgt gatgagggct ttgccgggtt ggactgcagc 1140
gagaagaggt gtcctgctga ctgtcacaat cgtggccgct gtgtagacgg gcggtgtgag 1200
tgtgatgatg gtttactggg agctgactgt ggggagctca agtgcctcaa tggctgcagt 1260
ggccatggcc gctgtgtcaa tgggcagtggt gtgtgtgatg agggctatac tggggaggac 1320
tgacagccagc tacggtgccc caatgactgt cacagtcggg gccgtgtgtg cgagggcaaa 1380
tgtgtatgtg agcaaggctt caagggctat gactgcagtg acatgagctg ccctaataac 1440
tgtcaccagc acggccgctg tgtgaatggc atgtgtgttt gtgatgacgg ctacacaggg 1500
gaagactgccc gggatgcgca atgccccagg cgtgcagca acaggggcct ctgtgtggac 1560
ggacagtgcg tctgtgagga cggcttcacc gggcctgact gtgcagaact ctctgtcca 1620
aatgactgcc atggccaggg tcgctgtgtg aatgggcagt gcgtgtgcca tgaaggattt 1680
atgggcaaa actgcaagga gcaaagatgt cccagtgact gtcattggcc gggccgctgc 1740
gtggacggcc agtgcattct ccacgagggc ttcacaggcc tggactgtgg ccagcactcc 1800
tgccccagtg actgcaacaa cttaggacaa tgcgtctcgg gccgtgtcat ctgcaacgag 1860
ggctacagcg gagaagactg ctgagaggtg tctcctccca aagacctcgt tgtgacagaa 1920
gtgacgggag agacgggtcaa cctggcctgg gacaatgaga tgcgggtcac agagtacctt 1980
gtcgtgtaca cgccaccca cgagggtggt ctggaaatgc agttccgtgt gcctggggac 2040
cagacgtcca ccatcatcca ggagctggaa cctggtgtgg agtactttat ccgtgtattt 2100
gccatcctgg agaacaagaa gagcattcct gtcagcgcca ggggtggccc gtacttacct 2160
gcacctgaag gcctgaaatt caagtccatc aaggagacat ctgtggaagt ggagtgggat 2220
cctctagaca ttgcttttga aacctgggag atcatcttcc ggaatatgaa taaagaagat 2280
gagggagaga tcacaaaaag cctgaggagg ccagagacct cttaccggca aactggtcta 2340
gctcctgggc aagagtatga gatattctct cacatagtga aaaacaatac ccggggccct 2400
ggcctgaaga gggtgaccac cacacgcttg gatgccccca gccagatcga ggtgaaagat 2460
gtcacagaca cactgcctt gatcacctgg ttcaagcccc tggctgagat cgatggcatt 2520
gagctgacct acggcatcaa agacgtgcca ggagaccgta ccaccatcga tctcacagag 2580
gacgagaacc agtactccat cgggaacctg aagcctgaca ctgagtacga ggtgtccctc 2640
atctcccgca gaggtgacat gtcaagcaac ccagccaaag agacctcac aacaggcctc 2700
gatgctccca ggaatcttcg acgtgtttcc cagacagata acagcatcac cctggaatgg 2760
aggaatggca aggcagctat tgacagttac agaattaagt atgcccccat ctctggaggg 2820
gaccacgctg aggttgatgt tccaaagagc caacaagcca caacaaaaac cacactcaca 2880
ggtctgaggg cgggaactga atatgggatt ggagtttctg ctgtgaagga agacaaggag 2940
agcaatccag cgaccatcaa cgcagccaca gagtgggaca cgcccaagga ccttcagggt 3000
tctgaaactg cagagaccag cctggaaga cctctggaag caccgttggc caaatctgac 3060
cgctaccgcc tcaattacag tctccccaca ggccagtggg tgggagtgca gcttccaaga 3120
aacaccactt cctatgtcct gagaggcctg gaaccaggac aggagtacaa tgtcctcctg 3180
acagccgaga aaggcagaca caagagcaag cccgcacgtg tgaaggcata cactgccatg 3240
ggctcccaa aggaagtcat tttctcagac atcactgaaa attcggctac tgtcagctgg 3300

```

```

agggcaccca cagcccaagt ggagagcttc cggattacct atgtgcccac tacaggaggt 3360
acaccctcca tggtaactgt ggacggaacc aagactcaga ccaggctggg gaaactcata 3420
cctggcgtgg agtaccttgt cagcatcatc gccatgaagg gctttgagga aagtgaacct 3480
gtctcagggg cattcaccac agctctggat ggcccatctg gcctggtgac agccaacatc 3540
actgactcag aagccttggc caggtggcag ccagccattg ccactgtgga cagttatgtc 3600
atctcctaca caggcgagaa agtgccagaa attacacgca cgggtgtccg gaacacagt 3660
gagtatgctc tgaccgacct cgagcctgcc acggaatata cactgagaat ctttgagag 3720
aaagggcccc agaagagctc aaccatcact gccaaagtta caacagacct cgattctcca 3780
agagacttga ctgctactga gggtcagtcg gaaactgccc tccttacctg gcgaccccc 3840
cgggcatcag tcaccgggta cctgctggtc tatgaatcag tggatggcac agtcaaggaa 3900
gtcattgtgg gtccagatac cacctcctac agcctggcag acctgagccc atccaccac 3960
tacacagcca agatccaggc actcaatggg ccctgagga gcaatatgat ccagaccatc 4020
ttcaccacaa ttggactcct gtaccccttc cccaaggact gctcccaagc aatgctgaat 4080
ggagacacga cctctggcct ctacaccatt tatctgaatg gtgataaggc tcaggcgctg 4140
gaagtcttct gtgacatgac ctctgatggg ggtggatgga ttgtgttcct gagacgcaaa 4200
aacggacggc agaacttcta ccaaactgg aaggcatatg ctgctggatt tggggacggc 4260
agagaagaat tctggccttg gctggacaac ctgaacaaaa tcacagccca ggggcagtac 4320
gagctccggg tggacctgcg ggaccatggg gagacagcct ttgctgtcta tgacaagttc 4380
agcgtgggag atgccaagac tcgctacaag ctgaagggtg aggggtacag tgggacagca 4440
ggtgactcca tggcctacca caatggcaga tccttctcca ctttgacaa ggacacagat 4500
tcagccatca ccaactgtgc tctgtcctac aaaggggctt tctggtacag gaactgtcac 4560
cgtgtcaacc tgatgggag atatggggac aataaccaca gtcagggcgt taactggttc 4620
cactggaagg gccacgaaca ctcaatccag tttgctgaga tgaagctgag accaagcaac 4680
ttcagaaatc ttgaaggcag gcgcaaacgg gcataaattc cagggaccac tgggtgagag 4740
aggaataagg ccagagcgca ggaaaggatt ttaccaaagc atcaatacaa ccagcccaac 4800
catcggtcca cacctgggca tttggtgaga gtcaaagctg accatggatc cctggggcca 4860
acggcaacag catgggcctc acctcctctg tgatttcttt ttcagcaaaa atctccagt gacaacatcg 4920
tctccaacat gtttctgttt taggtggctc tgggaatggg agaggggtag gatgtacagg 5040
caatagtttt ttacttctct gccgtatttt acatgaagct gtataattaa ttgtcattat 5100
ggtagtttgt tttagaacca tgtgtcattg gaagccatcc ctttttttac atttcatata 5160
ttttgttagc aaagattaaa actgtttcca ttttaaggat atgattaata ttattaatat 5220
acagaaacca gaaaagcaat tgaaaactaa ggatttttca agagatcttt ctttccaaaa 5280
aataatgatg atgatgatga ttgtattttt ttttttaaat aaaagcacia gtacttttga 5340
catttctgga cagtacctga
gtttgttaaa aac

```

<210> 66

<211> 3126

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503666CB1

<400> 66

```

attctggggc tcgggggac cgggacaccc tctcagctcc tgcccggggg cccatgtagt 60
cccttctgcc ctgtgcctcg gtgcctgtga cctgagcccc ttggttgacc ctgcactcgt 120
ccaacttggg ccaaacgact gccctcctt ctggcagtg gctggaccag ccggccagcg 180
ggagccccct tggcagaagc cggctcgtaa ggatcataaa ctggcggcgt ctggctgggg 240
cgaaggctgc tgaggtagga actgcgccag tcctagacgc cagaccgcgt cagaccctcc 300
tgccaggtga cagccgcaa gatggggctc tgggcctgc tgtggcctcc cctgtgttc 360
accgggctgc tcgtccgacc ccgggggacc atggccagg ccagtagctg ctctgtgaac 420
aaggacatct ttgaagtaga ggagaacaca aatgtcaccc agccgctggg ggacatccac 480
gtcccggagg gccaggagg gacctcgga gcctgtcca cccctttgc atttcggatc 540
cagggaaacc agctgtttct caacgtgact cctgattacg aggagaagtc actgcttgag 600
gtcagctgct tgtgtcagag cggaggcaca ttggtgacct agctaagggt gttcgtgtca 660
gtgctggacg tcaatgacaa tgcccccgaa ttccccctta agaccaagga gataagggtg 720
gaggaggaca cgaaaagtga ctccaccgtc atccccgaga cgcaactgca ggctgaggac 780
cgcgacaagg acgacattct gttctacacc ctccaggaaa tgacagcagg tgccagtgc 840
tacttctccc tggtagtgt aaaccgtccc gcctgaggc tggaccggcc cctggacttc 900
tacgagcggc cgaacatgac cttctggctg ctgggtgcggg acactccggg ggagaatgtg 960
gaaccagcc acactgccac cgccacacta gtgctgaacg tgggtgcccc cgacctgcgg 1020

```

```

ccccggtggt tccctgccttg cacctttctca gatgggtacg tctgcattca agctcagtag 1080
cacgggggctg tccccacggg gcacatactg ccatctcccc tcgtcctgcg tcccggaccc 1140
atctacgctg aggacggaga ccgcggcac tcattctacag catctttagg 1200
gactcggggca accctaccgt ggccaggagt gtccccagcc ccatgacctt ccttctgctg 1260
gtgaaggggca aacaggccga ccttgccccg tactcagtga cccaggtcac cgtggagggt 1320
gtggctgctg ccgggagccc gccccgcttc cccagagac tgtatcgtgg caccgtggcg 1380
cgtggcgctg gagcgggcgt tgtgggtcaag gatgcagctg ccccttctca gcctctgagg 1440
atccaggctc aggaccggga gttctcggac ctcaactcgg ccatcacata tcgaattacc 1500
aaccactcac acttcggat ggaggagag gtgtgtctga ccaccaccac actggcacag 1560
gcgggagcct tctacgcaga ggttgaggcc cacaacacgg tgacctctgg caccgcaacc 1620
acagtcattg agatacaagt ttccgaacag gagccccctt ccacagatgt ccccccattc 1680
ccagaggctg gaggaacaac tgggcccctg accagacca cttccgaggt cccagacccc 1740
cctgagccct cccagggacc ctccacgacc agctctgggg gaggcacagg cctcatcca 1800
ccctctggca caactctgag gccaccaaac tcgtccacac ccggggggcc cccgggtgca 1860
gaaaacagca cctcccacca accagccact cccggtgggg acacagcaca gaccccaag 1920
ccaggaacct ctacgccgat gcccccggt gtgggaacca gcacctcca ccaaccaggc 1980
acaccagtg ggggcacagc acagaccca gagccaggaa cctctcagcc gatgcccccc 2040
agtatgggaa ccagcacctc ccaccaacca gccacacccg gtggggggcac agcacagacc 2100
ccagaggcag gaacctctca gccgatgccc cccggtatgg gaaccagcac ctcccaccaa 2160
ccaaccacac ccggtggggg cacagcacag accccagagc caggaaacct tcagccgatg 2220
cccctcagca agagaccccc atcttcagggt ggcggccctt cggaggacaa gcgcttctcg 2280
ctgggtcgata tggcgccctt gggcggggtg ctgggtgcgc tgctgctgct ggctctcctt 2340
ggcctcgccg tccttgtcca caagcactat ggcccccgcc tcaagtgtg ctctggcaaa 2400
gctccggagc cccagcccca aggttttgac aaccaggcgt tcctccctga ccacaaggcc 2460
aactgggcgc ccgtccccag cccacgcac gacccaagc ccgcgaggc accgatgccc 2520
gcagagcccg cccccccg ccctgcctcc ccaggcgtg cccctgagcc ccccgagcg 2580
gcccagctg gcggaagccc cacggcggtg aggtccatcc tgaccaagga gcggcgcca 2640
gaggcggtg acaaggctgt ctggtttggc gaggacatcg ggacggaggc agacgtggtc 2700
gttctcaacg cgcccacctt ggaagtggtt ggcgccagtg actccggcag cggcgatgag 2760
ggcgagggcg cggggagggg tgggggtccc tacgatgcgc ccggtggtga tgactcctac 2820
atctaagtgg cccctccacc ctctcccca gccgcacggg cactggaggt ctgctcccc 2880
cagcctccga cccgaggcag aataaagcaa ggctcccgaa accaggcca tggcgtgggg 2940
caggcgcgcg ggtccatggg ggtcccattc actcagtcct ctgtcgtcat tagcgcttga 3000
gcccaggtgt gcagatgagg cggtgggtct ggcacgctg tccccaccc aaggctgcag 3060
cacttccgct aaaccacctg cagtgcgccg cgcttcccg aggtctgtg ccagctagtc 3120
tgggaa 3126

```

<210> 67

<211> 3066

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503668CB1

<400> 67

```

attctggggc tcgggggac ccggacaccc tctcagctcc tgccccgggg cccatgtagt 60
cccttctgcc ctgtgcctcg gtgcctgtga cctgagcccc ttggttgacc ctgcactcgt 120
ccaacttggg ccaaacgact gcccctcctt ctggcagtg gctggaccag ccggccagcg 180
ggagccccct tggcagaagc cggtcgtaaa ggatcataaa ctggcggcgt ctggctgggg 240
cgaaggtcgc tgaggtagga actgcgccag tcctagacgc cagaccgct cagaccctcc 300
tgccaggtga cagccgcca gatggggtct tgggcccctg tgtggcctcc cctgtgttc 360
accgggctgc tcgtccgacc cccgggggacc atggcccagg cccagtactg ctctgtgaac 420
aaggacatct ttgaagtaga ggagaacaca aatgtcaccg agccgctggt ggacatccac 480
gtcccggagg gccaggaggt gacctcgga gccttgcca cccctttg ccttctgag 540
cagggaacc agctgtttct caacgtgact cctgattacg aggagaagtc actgctttag 600
gctcagctgc tgtgtcagag cggaggcaca ttggacacga aagtgaactc caccgtcatc 660
ccgagagcgc aactgcaggc tgaggaccgc gacaaggacg acattctgtt ctacaccctc 720
caggaaatga cagcaggtgc cagtgactac ttctccctgg tgagtgtaaa ccgtcccgcc 780
ctgaggctgg accggccctt ggacttctac gagcgccga acatgacctt ctggctgctg 840
gtgctgggaca ctccggggga gaatgtggaa cccagccaca ctgccaccgc cactagtg 900
ctgaacgtgg tgccgcgcga cctgcggccc ccgtggttcc tgccctgcac cttctcagat 960

```



```

ggctacgtct gcattcaagc tcagtaaccac ggggctgtcc ccacggggca catactgcc 1020
tctcccctcg tcctgcgtcc cggacccatc tacgttgagg acggagaccg cggcatcaac 1080
cagcccatca tctacagcat ctttagggga aacgtgaatg gtacattcat catccacca 1140
gactcggggca acctcacgtg ggccaggagt gtcccagcc ccatgacctt ccttctgtg 1200
gtgaagggcc aacaggccga ccttgccgc tactcagtga ccaggtcac cgtggaggct 1260
gtggctgcgg ccgggagccc gcccgccttc cccagagac tgtatcgtgg caccgtggcg 1320
cgtggcgctg gagcgggctg tgtggtaag gatgcagctg ccccttctca gcctctgagg 1380
atccaggctc aggaccgga gttctcggac ctcaactcgg ccatacacata tcgaattacc 1440
aaccactcac acttcggat ggaggagag gttgtgctga ccaccaccac actggcacag 1500
gcgggagcct tctacgcaga gggtgaggcc cacaacacgg tgacctctgg caccgcaacc 1560
acagtcattg agatacaagt ttccgaacag gagccccct ccacagatgt ccccccattc 1620
ccagaggctg gaggaacaac tgggccctgg accagacca cttccgaggt cccagaccc 1680
cctgagccct cccagggacc ctccacgacc agctctgggg gaggcacagg cctcatcca 1740
ccctctggca caactctgag gccaccaacc tcgtccacac ccggggggcc cccgggtgca 1800
gaaaacagca cctccacca accgccact ccggtgggg acacagcaca gacccaaag 1860
ccaggaacct ctacgcgat ccccccggt gtgggaacca gcacctcca ccaaccagc 1920
acaccagtg ggggcacagc acagaccca gagccaggaa cctctcagcc gatgcccc 1980
agtatgggaa ccagcacctc ccaccaacca gccacaccg gtggggggcac agcacagacc 2040
ccagaggcag gaacctctca gccgatgccc cccggtatgg gaaccagcac ctcccacca 2100
ccaaccacac ccggtggggg cacagcacag accccagagc caggaaacct tcagccgatg 2160
cccctcagca agagacccc atcttcagg ggcgccctc cggaggacaa gcgcttctcg 2220
gtgggtggata tggcgccct gggcggggtg ctgggtgcgc tegtgtgtgt ggctctcct 2280
ggcctcgccg tccttgtcca caagcactat ggccccggc tcaagtgtgt ctctggcaa 2340
gctccggagc cccagcccca aggttttgac aaccaggcgt tcctccctga ccacaaggcc 2400
aactggcgcg ccgtccccag cccacgcac gaccccaag ccgcgagggc accgatgcc 2460
gcagagcccc cccccccg cctgcctcc ccaggcgtg cccctgagcc ccccgagcg 2520
gcccagagtg gcggaagccc cacggcggtg aggtccatcc tgaccaagga gcggcgggcc 2580
gagggcggtt acaaggctgt ctggtttggc gaggacatcg ggacggaggc agacgtggtc 2640
gttctcaacg cgccaccct ggacgtggat ggcccgagt actccggcag cggcgatgag 2700
ggcgagggcg cggggagggg tgggggtccc tacgatgcgc ccggtggtga tgactcctac 2760
atctaagtgg cccctccacc ctctcccca gccgcacgg cactggaggt ctgctcccc 2820
cagcctccga cccgaggcag aataaagcaa ggctccgaa accaggcca tggcgtgggg 2880
caggcgcgcg ggtccatggg ggtccattc actcagtcct ctgtcgtcat tagcgcttga 2940
gcccaggtgt cagatgagg cggtgggtct ggcacgctg tccccaccc aaggctgca 3000
cacttccgt aaaccacct cagtgcgcgc cgccttccg aggtctgtg ccagctagtc 3060
tgggaa 3066

```

<210> 68

<211> 3045

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7503672CB1

<400> 68

```

attctggggc tcgggggatc ccggacacc tctcagctcc tgccccgggg cccatgtagt 60
cccttctgcc ctgtgcctcg gtgcctgtga cctgagcccc ttggttgacc ctgcactcgt 120
ccaacttggg ccaaacgact gcccctcctt ctggcagtg gctggaccag ccggccagcg 180
ggagccccct tggcagaagc cggtcgtaaa ggatcataaa ctggcggcgt ctggctgggg 240
cgaaggtcgc tgaggtagga actgcgccag tcctagacgc cagaccgcct cagaccctcc 300
tgccaggtga cagccgcaa gatggggtct tggccctgc tgtggcctcc cctgctgttc 360
accgggctgc tcgtccgacc cccggggacc atggcccagg ccagtagctg ctctgtgaac 420
aaggacatct ttgaagtaga ggagaacaca aatgtcacc agccgctggt ggacatccac 480
gtcccggagg gccaggaggt gacctcgga gccttgtcca cccctttgc atttcggatc 540
cagggaaacc agctgtttct caacgtgact cctgattacg aggagaagtc actgcttgag 600
gctcagctgc tgtgtcagag cggaggcaca ttgtgacct agctaagggt gttcgtgtca 660
gtgctggacg tcaatgaca tgccccgaa ttcccttta agaccaagga gataagggtg 720
gaggaggaca cgaaagtga ctcaccgct atccccgaga cgcaactgca ggctgaggac 780
cgcgacaagg acgacattct gttctacacc ctccaggaaa tgacagcagg tgccagtgc 840
tacttctccc tggtagtggt aaaccgtccc gccctgaggc tggaccggcc cctggacttc 900
tacgagcggc cgaacatgac cttctggctg ctggtgcggg acactccggg ggagaatgtg 960

```

```

gaaccagccc acactgccac cgccacacta gtgctgaacg tgggtgcccgc cgacctgcgg 1020
cccccggtgt tccctgccctg cacccttctca gatggctacg tctgcattca agctcagtag 1080
cacgggggctg tccccacggg gcacatactg ccatctcccc tctgcctgcg tcccggaacc 1140
atctacgctg aggacggaga ccgcggcac cagactcgg gcaacctcac cgtggccagg 1260
ggaaacgtga atggtacatt catcatccac ctggtgaagg gccaacaggc cgaccttgcc 1320
agtgtcccca gcccacatgac cttccttctg ctggtgaagg gccaacaggc cgaccttgcc 1380
cgctactcag tgaccacagg tggcaccgtg gcgctggtg ctggagcggg cgttgtggtc 1440
ttccccccaga gactgtatcg tggcaccgtg gcgctggtg ctggagcggg cgttgtggtc 1500
aaggatgcag ctgccccctt cagcctctg aggatccagg ctgaggacc ggagtctctc 1560
gacctcaact cggccatcac atatcgaatt accaaccact cacacttccg gatggaggga 1620
gaggttgtgc tgaccaccac cacactggca caggcgggag ccttctacgc agaggttgag 1680
gcccacaaca cggtagacct tggcaccgca accacagtca ttgagataca agtttccgaa 1740
caggagcccc cctccacaga tgtccccca tccccagagg ctggaggaa aactgggccc 1800
tggaccagca ccacttccga ggtccccaga cccctgagc cctcccaggg accctccacg 1860
accagctctg ggggaggcac aggcctctg ccacctctg gcacaactct gaggccacca 1920
acctcgtcca caccggggg gccccgggt gcagaaaaca gcacctcca ccaaccagcc 1980
actcccggtg gggacacagc acagaccca aagccaggaa cctctcagcc gatgcccccc 2040
ggtgtgggaa ccagcacctc ccaccaacca gccacacca gtgggggac agcacagacc 2100
ccagagccag gaacctctca gccgatgccc cccagtatgg gaaccagca cccccacca 2160
ccagccacac ccggtggggg cacagcacag accccagagg caggaaacctc tcagccgatg 2220
ccccccggtg gcggccctc ggaggacaag cgcttctcg tggtgatat ggccggccctg 2280
ggcggggtgc tgggtgcgt gctgctgctg gctctcctg gcctcgccgt ccttgtccac 2340
aagcactatg gccccggct caagtgtgc tctggcaaag ctccggagcc ccagcccaa 2400
ggctttgaca accaggcgtt cctccctgac cacaaggcca actgggcgc cgtccccagc 2460
cccacgcacg accccaagcc cgcggaggca ccgatgccc cagagccgc cccagcctg 2520
cctgcctccc caggcgggtg ccttgagccc cccgcagcgg cccgagctgg cgggaagcccc 2580
acggcgggtg ggtccatcct gaccaaggag cggcgccag agggcgggta caaggctgtc 2640
tggttggcg aggcacatcg gacggaggca gacgtgtcg ttctcaacgc gccaccctg 2700
gacgtggatg gcgccagtga ctccggcagc ggcgatgagg gcgagggcgc ggggaggggg 2760
gggggtccct acgatgcgc cgggtggtgat gactcctaca tctaagtggc cctccaccc 2820
tctccccag ccgcacgggc actggaggtc tgcctcccc agcctccgac ccgaggcaga 2880
ataaagcaag gctcccgaaa cccaggccat ggcgtggggc aggcgcgcgg gtccatgggg 2940
gtcccattca ctcagtcctc tgtcgtcatt agcgcttgag cccaggtgtg cagatgaggc 3000
ggtgggtctg gccacgctgt cccacccca agcgtgcagc acttcccgta aaccacctgc 3045
agtgcgccgc gccttccga ggctctgtgc cagctagtct gggaa

```

<210> 69

<211> 4170

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 6039650CB1

<400> 69

```

caagatgatg tgcgaggatg tgccgaccat cagcgaagca gaaggccccc ctggaggagg 60
tggaggccat ggttccggct ccccttcaca gccagatgca gattcacatt ttgaacagtt 120
gatggtctcc atgctagaag aaagggaccg ccttcttgat acactgagag agactcaaga 180
aacgctggcc ttaaccacagg ggaagttaca cgaggttggg catgaaagag attccttgca 240
gagacagctc aacacggcac ttccacagga gttcgcagca cttactaaag aactcaatgt 300
atgcagggaa cagctccttg aaagggaaga agaaattgct gaactgaaag cagaaaggaa 360
taacaccagg ctgctgttag agcatttgga atgccttgct tccaggcatg agcggctctc 420
taggatgacc gtggtgaaga gacaagcgca gtctccagca ggcgtgtcca gcgaagtggg 480
agtgtgaaa gcaactgaagt ccttatattga acaccacaaa gctctggatg aaaaggtgag 540
agagcgatta cgagtagcac ttgaaagatg tagtttggtt gaagagggaat taggtgccac 600
acacaaagag ctaatgattc ttaaagaaca gaataatcag aaaaaaactc taacagatgg 660
agtgtggagc ataaaccatg aacaagaaaa tacaccaagc acgagtggaa agagatcttc 720
tgatggttct ttaagccacg aggaagacct tgctaaagta attgagctcc aggaatcat 780
aagtaagcag tcaagggaac agagccaaat gaaagaacgc ctggcttccc tttccagtca 840
tgtgacagaa ctggaagagg atctggacac ggctagaaaa gatctcatca aatctgaaga 900
aatgaacaca aaattgcaac gagatgtccg tgaagccatg gcccaaaagg aagatatgga 960
agagagaatc actactcttg aaaaacgcta cctcgtgca cagcgtgaag ccacatctgt 1020

```

```

gcatgacctc aatgataaac ttgaaaatga aattgcaaat aaagattcta tgcacgcaca 1080
gactgaagat aaaaaccgcc agttacagga gcgcttgga ttggcagagc aaaagctgca 1140
acagacactg aggaaggcag agacgctccc ggaggtggag gcggagctgg ccagaggggt 1200
ggcagcgctt tccaaggctg aagagagaca cggcaacatt gaagaaaggt tacgacagat 1260
ggaagcacag ttggaggaga agaatcgaga actgcagcgg gcaaggcaaa gagaaaaaat 1320
gaacgaagaa cataataaac gtttatcaga cactgttgac aagctgcttt cagaatctaa 1380
tgagaggctt caacttcac ttaaagagag aatggctgct ttggaagata agaactctct 1440
tttaagagaa gttgaaagtg caaaaaagca gttagaagaa acacaacacg ataaggatca 1500
gcttgtctca aacattgaag cactgagggc tgaactagac cacatgagac taagaggtgc 1560
ttcacttcat catggccgac ccacttggg cagtgtccca gatttcaggt tccccatggc 1620
agacggccac acagactcct acagcaccag tgcagtgtcg cggcgccac agaaaggccg 1680
gctggcagcc ctgagagatg agccttccaa ggtacaaact cttaatgagc aggattggga 1740
acgtgccag caagctagtg tcttggcaaa tgtagcacia gcattcgaga gtgatgctga 1800
cgtgtctgat ggtgaagatg acagggacac tctctcagc tcagttgacc tgctatcgcc 1860
cagcgggcag gccagcgcgc acacactagc catgatgctt caggagcagc tggacgccat 1920
caacaaagag atcaggttga ttcaggaaga aaaagaaaat acagagcagc gggcagagga 1980
gattgaaagt cgagttggca gtggaagtct agacaatctt ggtcgtttta gatcaatgag 2040
ctccattccc ccctaccctg ctctctcgct tgctagctcc tccctccgg gcagtgggcg 2100
ctccacccca cgaaggatcc ctacagccc agctcgggaa gtggacagac tgggcgtcat 2160
gacccttttg ccaccttcca gagaagaggt acgagatgac aagacaacca taaagtgtga 2220
aacctccccg ccttctctcc cgagagccct tcggttagac cggctgcaca agggggcgct 2280
gcacaccctc agccacgagg acatcaggga cataaggaaac tccacaggct ccaggatgg 2340
tcccgtagac aacccagca gtagcaacag tagccaggac tcgctccaca aagcccaaaa 2400
gaagaaaggc attaagtcct ccattggccg cttgtttggc aagaaagaaa agggccgacc 2460
tggaacaaact ggcaaagaag cattaggaca agctggtgtt tccgagacgg ataactcatc 2520
tcaggatgcc ttgggactta gcaaattggg gggacaggct gaaaaaaatc gtaaacttca 2580
aaaaaagcat gaattgctgg aggaagcccg gagacaaggt ttaccttttg ccaatggga 2640
cgggccaaacg gttgtgtct ggtgtgtggt atgccagcct ggtatgtggc 2700
tgcttgcga gcaaacgtga aaagcggggc catcatgtcg gccctgtccg acacagagat 2760
ccagcgtgag attggcatca gcaacccct gcacaggctg aagctgaggc tggccatcca 2820
ggagatcatg tcgctgacca gccgctctgc cccgccaca tctagaacga ccacaggaaa 2880
tgtctggtta acacacgaag agatggaaac gctcgcagcc acgccgcaaa cggaagatga 2940
ggaggggaagc tgggctcaga cactcgccta tggggacatg aaccacgagt ggatcggcaa 3000
cgagtggctc cccagcctgg gctccccca gtaccgcagc tacttcatgg agtcgctgt 3060
agacgccagg atgctggacc acctgaccaa gaaagacctt cgagggcagc tgaaaatggt 3120
cgacagtttt cacagaaaca gtttccagtg tggaattatg tgcctgagaa ggttaaatta 3180
tgaccggaag gaactggaaa gaaaaagaga agaaagtcag agtgaaataa aagacgtgct 3240
tgtttgagac aatgatcgag tgattcgctg gatcctgtca attggcctta aagaatatgc 3300
aaacaatctt atagagagtg gtgttcacgg agcacttctg gccttagatg aaaccttca 3360
cttcagtgc tgggactgc gtgtacagat cccgacgcag aacacacagg ctctgtctgt 3420
cttgaaaga gaatttaaca accttttggg catggggact gatagaaggt ttgatgaaga 3480
tgatgataaa agcttttaga gagcaccttc atggagaaaa aagtttagac caaggacat 3540
tcgtggctta gctgctgggt cagcagagac tctccctgca aacttccggg tgacttcttc 3600
tatgtcttcc cctctatgc agccaaagaa gatgcagatg gacggcaatg tatcaggaac 3660
acagaggttg gattctgcta cagtccaggac ttactcctgc taaagtctcc tgttgtttac 3720
ccactact tctacagatg attatgcagc attgaaatcc acataagact acattttaga 3780
atccagtgga atctttaatc ttgttaatac ttgttatatg gacctaaaga tattttatta 3840
cagagttttt aattagtga aaattcatga ataccataga gaaaatatat tagaatttaa 3900
tgtttcttat atttatgtaa acttatgact cttcatttat atagttactt actttttcat 3960
gtatatccag gctataaata tcttttcaaa tcatgttctt atacctaatt ttagtctttc 4020
aaatgaatgt actgtaatgc ttgtatgtat aaatcctatg aatagagggc ttttgtaaat 4080
tatgcattta ttgtaattat cattaatttt ttaatgataa accatgacaa aggattttac 4140
gtttataaaa ttatgacaga agccatgtgc 4170

```

<210> 70

<211> 3092

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7509919CB1

<400> 70

```

agacgtgacc cagggcagac tggtagcaaa gccccacgc ccagccagga gcaccgccga 60
ggactccagc acaccgaggg acatgctggg cctgcgcccc ccactgctcg ccctgggtggg 120
gctgctctcc ctccgggtgcg tcctctctca ggagtgcacg aagttcaagg tcagcagctg 180
ccgggaatgc atcgagtcgg gggccggctg cacctggtgc cagaagctgg ctgcacctac 240
gctcagccgt ctccctgggc ttcttgaatc tggacaagtc ctacagcttc tttgtctttc 300
ctgtccttgg gcgtttctga ggaagtcagg agagcagctt tgtagcatga ccctcggggg 360
tcttgtctgg agtttcctct gcctggactc agggagcggc tttgggtcct ccagcccacg 420
ccaccgcggc gagcagtcct cggaggagaa gccgggtggc tcccgctctg cctcggtactg 480
agcagggtgg ccaggcccca gaacttcaca gggccggggg atcctgactc cattcgctgc 540
gacacccggc cacagctgct catgaggggc tgtgcggctg acgacatcat ggaccccaca 600
agcctcgctg aaaccagga agaccacaat gggggccaga agcagctgtc cccacaaaaa 660
gtgacgcttt acctgcgacc aggccaggca gcagcgttca acgtgacctt cgggcggggc 720
aagggtacc ccacgacct gtactatctg atggacctct cctactccat gcttgatgac 780
ctcaggaatg tcaagaagct aggtggcgac ctgctccggg cctcaacga gatcaccgag 840
tccggccgca ttggcttcgg gtccttcgtg gacaagaccg tgctgcggtt cgtgaacacg 900
caccctgata agctgcgaaa cccatgcccc aacaaggaga aagagtgcc a gccccggtt 960
gccttcaggc acgtgctgaa gctgaccaac aactccaacc agtttcagac cgagggtcggg 1020
aagcagctga tttccgaaa cctggatgca cccgagggtg ggctggacgc catgatgcag 1080
gtcgcgcctt gcccggagga aatcggtctg cgcaacgtca cgcggctgct ggtgtttgcc 1140
actgatgacg gcttccattt cgcgggcgac gggaagctgg gcgccatcct gacccccaac 1200
gacggccgct gtcacatgga ggacaacttg tacaagagga gcaacgaatt cgataccaca 1260
tcgggtgggccc agctggcgca caagctggct gaaaacaaca tccagcccat cttcgcggtg 1320
accagtagga tggatgaagac ctacgagaaa ctaccgaga tcatcccaa gtcagccgtg 1380
ggggagctgt ctgaggactc cagcaatgtg gtccatctca ttaagaatgc ttacaataaa 1440
ctctcctcca gggcttctct gcatcacaac gccctccccg acacctgaa agtcacctac 1500
gactccttct gcagcaatgg agtgacgcac aggaaccagc ccagaggtga ctgtgatggc 1560
gtgcagatca atgtcccgat cacttccag gtgaaggcca cggccacaga gtgcacccag 1620
gagcagtcgt ttgtcatccg ggcgctgggc ttacaggaca tagtgaccgt gcaggtcctt 1680
ccccagtggt agtgccgggt cccggaccag agcagagacc gcagcctctg ccatggcaag 1740
ggcttcttgg agtgccggcat ctgcaggtgt gacactggct acattgggaa aaactgtgag 1800
tgccagacac agggccggag cagccaggag ctggaaggaa gctgccggaa ggacaacaac 1860
tccatcatct gctcagggct gggggactgt gtctgcgggc agtgccctgt ccacaccagc 1920
gagctccccc gcaagctgat atacgggcag tactgaggt gtagacccat caactgtgag 1980
cgctacaacg gccaggtctg cggcgccccc gggagggggc tctgcttctg cgggaagtgc 2040
cgctgccacc cgggctttga gggctcagcg tgccagtgcg agaggaccac tgagggctgc 2100
ctgaaccgca ggcgtgttga gtgtagtggt cgtggccggg gccgctgcaa cgtatgcgag 2160
tgccattcag gctaccagct gcctctgtgc caggagtgcc ccggtgccc ctcacctgt 2220
ggcaagtaca tctcctgcgc cgagtgcctg aagttcgaaa agggcccctt tgggaagaac 2280
tgcagcgcgg cgtgtccggg cctgcagctg tcgaacaacc cgtgaaggg caggacctgc 2340
aaggagaggg actcagaggg ctgctgggtg gcctacacgc tggagcagca ggacgggatg 2400
gaccgctacc tcatctatgt ggatgagagc cgagagtgtg tggcaggccc caacatcgcc 2460
gccatcgctg gggggcaccg tggcaggcat cgtgctgac ggcatctct ccattgtcat 2520
ctggaaggct ctgatccacc tgagcgacct cggggagtac aggcgctttg agaaggagaa 2580
gctcaagtcc cagtgaaca atgataatcc ctttttcaag agcgccacca cgacggtcat 2640
gaacccaag tttgttgaga gttaggagca ctgttggaag acaaggccgt caggaccac 2700
catgtctgcc ccatcacgcg gccgagacat ggcttgccac agctcttgag gatgtacca 2760
attaaccaga aatccagtta ttttccgccc tcaaatgac agccatggca ggccgggtgc 2820
ttctgggggc tcgtcggggg gacagctcca ctctgactgg cacagtctt gcattggagac 2880
ttgaggagg agggcttgag gttggtgagg ttaggtgcgt gtttctgtg caagttagga 2940
catcagctct attaaagggt gtgccaattt atttacattt aaacttgtca ggggtataaaa 3000
tgacatccca ttaattatat tgttaataaa tcacgtgtat agaaaaaaa taaaacttca 3060
atacaggctg tccatggaaa aaaaaaaaaa aa 3092

```

<210> 71

<211> 3257

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510758CB1

<400> 71

```

tcttagtcgc cgtctctctc ctcacctctc agggaaaggg ggggacatag gggcgctcgc 60
gggccccggc gaatgcgccc ccgcgcgcct ctcggtgtgc gccgcctcgc ggggatgaag 120
caccggccgt gaagatggag gtgacctgcc ttctacttct ggcgctgatc cccttccact 180
gccggggaca aggagtctac gctccagccc aggcgcagat cgtgcatgcg ggccaggcat 240
gtgtggtgaa agaggacaat atcagcgagc gtgtctacac catccgggag ggggacaccc 300
tcatgctgca gtgccttgta acagggcacc ctgcacccca ggtacggtgg accaagacgg 360
caggtagcgc ctcggaacaag ttccaggaga catcggtgtt caacgagacg ctgcgcatcg 420
agcgtattgc acgcacgcag ggcggccgct actactgcaa ggctgagaac ggcggtgggg 480
tgccggccat caagtccatc cgcgtggacg tgcagtacct ggatgagcca atgctgacgg 540
tgcaccagac ggtgagcgat gtgcgaggca acttctacca ggagaagacg gtgttcctgc 600
gctgtactgt caactccaac ccgcctgccc gcttcatctg gaagcggggt tccgataccc 660
tatccacacg ccaggacaat ggggttgaca tctatgagcc cctctacact cagggggaga 720
ccaaggtcct gaagctgaag aacctgcggc cccaggacta tgccagctac acctgccagg 780
tgtctgtgcg taacgtgtgc ggcattccca acaaggccat caccttccgg ctcaccaaca 840
ccacggcacc accagccctg aagctgtctg tgaacgaaac tctggtggtg aacctggggg 900
agaatgtgac ggtgcagtgt ctgctgacag gcggtgatcc cctccccag ctgcagtggg 960
cccatggggc tggcccactg cccctgggtg ctctggccca ggtgggcacc ctgagcatcc 1020
cttcagtgca ggcccgggac tctggctact acaactgcac agccaccaac aatgtgggca 1080
accctgccaa gaagactgtc aacctgctgg tgcgatccat gaagaacgct acattccaga 1140
tcaactctga cgtgatcaaa gagagtgaga acatccagct gggccaggac ctgaagctat 1200
cgtgcccggt ggatgcagtg cccaggaga aggtgacctt cagatgggtc aagaatggca 1260
agccggcacg catgtccaag cggctgctgg tgaccgcgca tgatcctgag ctgccgcgag 1320
tcaccagcag cctagagctc attgacctgc acttcagtga ctatggcacc tacctgtgca 1380
tggtctcttt cccaggggca cccgtgcccg acctcagcgt cgaggtcaac atctcctctg 1440
agacagtgcc gccaccatc agtgtgccc agggtagggc cgtggtgacc gtgcgcgagg 1500
gatcgctgca cgagctgcaa tgcgaggtgc ggggcaagcc gcggccgcca gtgctctggt 1560
cccgcgtgga caaggaggct gcactgctgc cctcggggct gccctggag gagactccgt 1620
acgggaagct gcggctggag cgagtgcgac gagacatgag cgggacctac cgctgccaga 1680
cggcccgcta taatggcttc aacgtgcgcc cccgtgaggc ccaggtgcag ctgaacgtgc 1740
agttcccgcc ggaggtggag cccagttccc aggacgtgcg ccaggcgctg ggcgggccc 1800
tgctcctgca ctgctcgctg ctgcgaggca gccccagcg catcgctcgt gctgtgtggc 1860
gtttcaaagg gcagctgctg ccgcgcgcgc ctggtgttcc cgcgcgcgcg gaggcgccgt 1920
atcacgcgga cgtgcgcctc gacgcgtaaa ctgcgcagc cagcggcagc tacgagtgca 1980
gcgtctccaa cgatgtgggc tgggtgcct gcctcttcca ggtctccgcc aaagcctaca 2040
gcccgaggtt ttacttgcag acccccaacc ccacccgcag ccacaagctg tccaagaact 2100
actcctacgt gctgcagtgg actcagaggg agcccgcgc tgtcgaccct gtgctcaact 2160
acagactcag catccgccag ttgaaccagc acaatgcggt ggtcaaggcc atcccggctc 2220
ggcgtgtgga gaaggggcag ctgctggagt acatcctgac cgatctccgt gtgccccaca 2280
gctatgaggt ccgcctcaca cctatacca ccttcggggc tggtgacatg gcctcccgca 2340
tcatccacta cacagagccc atcaactctc cgaacctttc agacaacacc tggcactttg 2400
aggatgagaa gatctgtggc tataccaggg acctgacaga caactttgac tggacgcggc 2460
agaatgccct caccagaac cccaaacgct cccccaacac tgggtcccccc accgacataa 2520
gtggcaccctc tgagggctac tacatgttca tgcgacatc gaggcctcgg gagctggggg 2580
accgtgcaag gttagttagt cccctctaca atgccagcgc caagttctac tgtgtctcct 2640
tcttctacca catgtacggg aaacacatcg caagagcaaa actccatttc aaaaaaaa 2700
aaggaaaaag aaaagactcc agcaggatga gatggaagtg gaggggatgg agaagacaac 2760
taagaggatg caaatatgga gtgactgcca ggtgatgtcc tgggtgcgcc gcatccttga 2820
ttgcattaac cctcacagcc accagatgag ggaagtgtta ccatccaact cagagacgtc 2880
aagcattttt cccagaccac acagtaagct agatctgtgt gacttctcca agccctaaac 2940
atgagatcgg aagggtgctt ctttgtgggc agtgagagga agtcctgctt cccaagttag 3000
ggagtgaaca taaggagtgc ccttgggac gcaatggtac tggtaggagc gtcacataac 3060
cagtgtcagc catcacttcc agagctcctg ctcagttcca ggtatcaggt gctgtagtga 3120
acgaggtgga cgagttcccc atcctcttgg catttacagg caagtgggag agacaggtat 3180
taaacaactc atgctaccag ccaggagtta gtgaccctg tggcagtgga aggacagggc 3240
tgagactggg ggtaaca

```

<210> 72

<211> 3360

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510063CB1

<400> 72

```

gggactgtca ctcatctctcc gatcagcgcg tgaacgcagc tcggctgccg ctggcaggga 60
aacaattctg caaaaataat catactcagc ctggcaattg tctgccccta ggtctgtcgc 120
tcagccgccg tccacactcg ctgcaggggg gggggcacag aatttaccgc ggcaagaaca 180
tccctcccag ccagcagatt acaatgctgc aaactaagga tctcatctgg actttgtttt 240
tccctgggaac tgcagtttct ctgcaggtgg atattgttcc cagccagggg gagatcagcg 300
ttggagagtc caaattcttc ttatgccaaag tggcaggaga tgccaaagat aaagacatct 360
cctgggttctc ccccaatgga gaaaagctca ccccaaacca gcagcggatc tcagtgggtg 420
ggaatgatga ttctcctcct accctcacca tctataacgc caacatcgac gacgccggca 480
tttacaagtg tgtggttaca ggcgaggatg gcagtgaagc agaggccacc gtcaacgtga 540
agatctttta gaagctcatg ttcaagaatg cgccaacccc acaggagttc cggggggggg 600
aagatgccgt gattgtgtgt gatgtggtga gctcccctcc accaaccatc atctggaaac 660
acaaaggccg agatgtcatc ctgaaaaaag atgtccgatt catagtctcg tccaacaact 720
acctgcagat ccggggcatc aagaaaacag atgagggcac ttatcgctgt gagggcagaa 780
tccctggcacg gggggagatc aacttcaagg acattcaggt cattgtgaat gtgccacct 840
ccatccaggc caggcagaat attgtgaatg ccaccgccaa cctcgccag tccgtcacc 900
tgggtgtgcga tgccgaacgg ttcccagagc ccaccatgag ctggacaaag gatggggaac 960
atagatagca agaggaagac gatgagaagt acatcttcag cgacgatagt tccagctga 1020
ccatcaaaaaa ggtgggataag aacgacaggt ctgagtacat ctgcattgct gagaacaagg 1080
ctggcgagca ggatgcgacc atccacctca aagtctttgc aaaacccaaa atcacatatg 1140
tagagaacca gactgccatg gaattagagg agcaggtcac tcttacctgt gaagcctccg 1200
gagaccccat tccctccatc acctggagga cttctacccg gaacatcagc agcgaagaaa 1260
aggtctcggt gactcgacca gagaagcaag agactctgga tgggcacatg gtggtgcgt 1320
gccatgcccc tgtgtcgtcg ctgacctga agagctcca gtacactgat gccggagagt 1380
acatctgcac cgccagcaac acctcggcc aggactcca gtccatgtac cttgaagtgc 1440
aatatgcccc aaagctacag ggccctgtgg ctgtgtacac ttggggagggg aaccaggtga 1500
acatcacctg cgaggtatth gcctatccca gtgccacgat ctcatgggtt cgggatggcc 1560
agctgctgcc aagctccaat tacagcaata tcaagatcta caacaccccc tctgccagct 1620
atctggaggt gaccccagac tctgagaatg attttgggaa ctacaactgt actgcagtga 1680
accgattgg gcaggagtcc ttggaattca tccctgttca agcagacacc cctcttccac 1740
catccatcga ccaggtggag ccatactcca gcacagccca ggtgcagttt gtgaaccag 1800
aggccacagg tggggtgccc atcctcaaat acaaagctga gtggagagca gtgggtgaag 1860
aagtatggca ttccaagtgg tatgatgcca aggaagccag catggagggc atcgccacca 1920
tcgtgggcct gaagcccga acaacgtacg ccgtaaggct ggccggcgctc aatggcaaa 1980
ggctgggtga gatcagcgcg gcctccgagt tcaagacgca gccagtccaa ggggaacca 2040
gtgcacctaa gctcgaaggg cagatgggag aggatggaaa ctctattaaa gtgaacctga 2100
tcaagcagga tgacggcgcg tcccccatca gacctatct ggtcaggtac cgagcgaa 2160
tctcctccga gtggaaacca gagatcaggc tcccgctctg cagtgaccac gtcatgctga 2220
agtccctgga ctggaatgct gagtatgagg tctacgtggt ggctgagaac cagcaaggaa 2280
aatccaaggc ggctcattht gtgttcagga cctcggccca gccacagcc atcccagcca 2340
acggcagccc cacctcaggc ctgagcaccg gggccatcgt gggcatcctc atcgatcat 2400
tcgtcctgct cctgggtggt gtggacatca cctgtactt cctgaacaag tgtggcctgt 2460
tcatgtgcat tgcggtcaac ctgtgtggaa aagccgggcc cggggccaag ggcaaggaca 2520
tggaggaggg caaggccgcc ttctcgaaag atgagtccaa ggagcccatc gtggaggttc 2580
gaacggagga ggagaggacc ccaaaccatg atggagggaa acacacagag cccaacgaga 2640
ccacgccact gacggagccc gagaagggcc ccgtagaagc aaagccagag tgccaggaga 2700
cagaaacgaa gccagcgcca gccgaagtca agacgggtccc caatgacgcc acacagacaa 2760
aggagaacga gagcaaagca tgatgggtga agagaaccga gcaaagatca aaataaaaa 2820
tgacacagca gcttcaccag agcatttcca acaccacaga cacacacacg cagcacaca 2880
cacaaacaca catgcacaca cacacatctc atttctctag tgtcttttgc ctttaaaaaa 2940
aactaaacag ataaaacatg ggaatctcct tttttagtgt ttattagaag agggcccatg 3000
ttgggggact ccctagtgtt agaaaaatgg ataacaatat gtgttaacct cagaagtc 3060
gctagggaca cgccggttcc ctgaagaccg ttaaaatcta tcctacgagg aagggcctcc 3120
aatattaagg aatctgaggg aagctcagga taccgaaatt caaggatcca gggaggccca 3180
cacattgggg gtgtcccccga ttgccacagg accaggtttc agggatattc ctttcaggcc 3240
actatagcta atcggattga ccaagtcccc agttttatta ttatacttgc agtcaagttg 3300
gaaccttgga aacctccaat attaggtata tgtccgggtc acttggtttt ggtcagtatg 3360

```

<210> 73

<211> 6871

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7510135CB1

<400> 73

```

gcgatggctt ttccgcccgcg gcgacggctg cgccctcggtc cccgcggcct cccgcttctt 60
ctctcggggac tcttgctacc tctgtgccgc gccttcaacc tagacgtgga cagtcctgcc 120
gagtactctg gccccgaggg aagttacttc ggcttcgcgc tggatttctt cgtgcccagc 180
gcgtcttccc ggatgtttct tctcgtggga gctcccaaag caaacaccac ccagcctggg 240
attgtggaag gagggcaggt cctcaaagt gactgggtct ctaccgcgcg gtgccagcca 300
attgaatttg atgcaacagg caatagagat tatgccaaag atgatccatt ggaatttaag 360
tcccacagtt ggtttggagc atctgtgagg tcgaaacagg ataaaatttt aacctgtgcc 420
ccattgtacc attggagaac tgagatgaaa caggagcgag agcctgttgg aacatgcttt 480
cttcaagatg gaacaaagac tgttgagtat gctccatgta gatcacaaga tattgatgct 540
gatggacagg gattttgtca aggaggattc agcattgatt ttactaaata aacattgctt 600
atttgaaata gggagacatc ccccatcact ctgggaatgg aggaaagtag acactctcct 660
cagtaaatga taggctgaca gactacttct tgggtggctc ggtagctttt attggcaagg 720
tcagcttatt tcggatcaag tggcagaaat gctatctaaa agcagcccca atgtttacag 780
catcaagtat aataaccaat tagcaactcg gactgcacaa gctatttttg atgacagcta 840
tttgggttat tctgtggctg tcggagattt caatgggtgat ggcatagatg actttgtttc 900
aggagttcca agagcagcaa ggactttggg aatggtagga cagttaaaag gtttatattt 960
atgatgggaa gaacatgtcc tccttataca attttactgg cgagcagatg gctgcatatt 1020
tcggattttc ttagctgccc actgacatta atggagatga ttatgcagat gtgtttattg 1080
gagcacctct cttcatggat cgtggctctg atggcaaac aaagctgaat ggaattgagg 1140
cagtgtctct acagagagct tcaggagact tccagacgac aaagctgaat ggatttgagg 1200
tctttgcacg gtttggcagt gccatagctc ctttgggaga tctggaccag gatggtttca 1260
atgatattgc aattgctgct ccatatgggg gtgaagataa aaaaggaatt gtttatatct 1320
tcaatggaag atcaacaggc ttgaacgcag tcccatctca aatccttgaa gggcagtggg 1380
ctgctcgaag catgccacca agctttggct attcaatgaa aggagccaca gatatagaca 1440
aaaatggata tccagactta attgtaggag cttttgggtg agatcgagct atcttataca 1500
gggccagacc agttatcact gtaaatgctg gtcttgaagt gtaccctagc attttaaatc 1560
aagacaataa aacctgctca ctgcctggaa cagctctcaa agtttctgtt tttaatgtta 1620
ggttctgctt aaaggcagat ggcaaaggag tacttcccag gaaacttaat ttccaggtgg 1680
aacttctttt ggataaactc aagcaaaaagg gagcaattcg acgagcactg tttctctaca 1740
gcaggtcccc aagtcactcc aagaacatga ctatttcaag ggggggactg atgcagtgtg 1800
aggaattgat agcgtatctg cgggatgaat ctgaatttag agacaaactc actccaatta 1860
ctatttttat ggaatatcgg ttggattata gaacagctgc tgatacaaca ggcttgcaac 1920
ccattcttaa ccagttcacg cctgctaaca ttagtcgaca ggctcacatt ctacttgact 1980
gtggtgaaga caatgtctgt aaaccacaagc tggaagtttc ttagatagat gatcaaaaga 2040
agatctatat tggggatgac aaccctctga cattgattgt taaggctcag aatcaaggag 2100
aagggtgcta cgaagctgag ctcatcgttt ccattccact gcaggctgat ttcatcgggg 2160
ttgtccgaaa caatgaagcc ttagcaagac ttctctgtgc atttaagaca gaaaacccaa 2220
ctcgccaggt ggtatgtgac cttggaaacc caatgaaggc tggaactcaa ctcttagctg 2280
gtcttcgttt cagtgtgcac cagcagtcag agatggatac ttctgtgaaa tttgacttac 2340
aaatccaaag ctcaaactca tttgacaaag taagcccagt tgtatctcac aaagttgatc 2400
ttgctgtttt agctgcagtt gagataagag gagtctcgag tctgatcat atctttcttc 2460
cgattccaaa ctgggagcac aaggagaacc ctgagactga agaagatgtt gggccagtgt 2520
ttcagcacat ctatgagctg agaaacaatg gtccaagttc attcagcaag gcaatgctcc 2580
atcttcagtg gccctacaaa tataataata acactctgtt gtatatcctt cattatgata 2640
ttgatggacc aatgaactgc acttcagata tggagatcaa ccctttgaga attaatctct 2700
catctttgca aacaactgaa aagaatgaca cgggtgccgg gcaagggtgag cgggaccatc 2760
tcatcactaa gcgggatctt gccctcagtg aaggagatat tcacactttg ggttgtggag 2820
ttgctcagtg cttgaagatt gtctgccaag ttgggagatt agacagagga aagagtgcga 2880
tcttgtacgt aaagtcatta ctgtggactg agacttttat gaataaagaa aatcagaatc 2940
attcctattc tctgaagtcg tctgcttcac ttaatgtcat agagtttctt tataagaatc 3000
ttccaattga ggatatcacc aactccacat tgggttaccac taatgtcacc tggggcattc 3060
agccagcgcc catgcctgtg cctgtgtggg tgatcatttt agcagttcta gcaggattgt 3120
tgctactggc tgttttggtt tttgtaatgt acaggatggg ctttttttaa cgggtccggc 3180
cacctcaaga agaacaagaa agggagcagc ttcaacctca tgaaaatggt gaaggaaact 3240
cagaaactta actgcagttt ttaagttatg ctacatcttg acccactaga attagcaact 3300

```

```

ttattataga tttaaacttt cttcatgagg agtaaaaaatc caaggccttta ctgctgatag 3360
tgctaattgg cattaaccac aaaatgagaa ttatatattgt caaccttctc cttataaata 3420
agttcagaca tacattttaat aacatagggt gacttggtgt tttagggtatt taaataataa 3480
aatttcaagg gatagttttt attcaatgta tataagacag gtagtgccctg atttactact 3540
ttatataaaa tagtacctcc ttcagttact gtttctgatt taatgtacgg aactttattt 3600
gttggtgttg ttgttggtgt tgttggtgtt ttaaagcagt ccaaatttgg accttagcaa 3660
tcatgtcttt tgtataggta cttaatgtta atacatatta cactacagtt tacttttcag 3720
aatactaaag actttataac tgcatagaact tggatttttt taatcactca tatggtagaa 3780
ttttataaac acatagatga taccatccaa attcctgctt ttaataacaa aggtacaata 3840
ttttgtttta ctatgaaaat ctggtagatc ctattacact tctgtttata ttaaattccac 3900
aatattttat tacattttta acttgtataa attttaggtc aaatccttca agccaacct 3960
tactaaaaat tagttccata atcacaaatg gctcttttgt gtaattgttt aatttcacct 4020
gaatatcata atgcttaaaag ccatatggag ttggaaaatta tttccaaagc atatttattc 4080
cattgtttta gtctggctat ttacagtata aaaaaagcat ttttattaaa atactgtgta 4140
gttctttgag atagtgtctt atgcatatag taagtattac attccttagag tagagcagaa 4200
tttttagtta gtatttaattt attttccctc attcatgtac ttttccttat atttccaaaa 4260
ctgttactga gaatgggtca agatcagtga gaaatcttta cagttgacag gaacctggac 4320
cccttaccct aactttatga gtaatgcttg gaattaaaac tcttaaggca actcactgat 4380
ttacttctag caatagcatg atgttacagg aatattacct ctgtttaagc aaggtaatgt 4440
gtaaaatcag tctcggtctg cagaataact tctaaaagggt atttttataa gcagttcaag 4500
ttactgaaaa ccttttaaac ctttctgaag ttcgttagta taaattactt tcttaggatt 4560
attaataaaa gccacatagg tggcaagttg tagttttata tggctctgta gagtgggtgaa 4620
ccttctagag gaatatatga tttattcaca gttcctcaag gcctggggat gatgatcagt 4680
tatacctatt tttgtgcaat tacatcatgt tgtacattag aaatggagag tttaatagct 4740
ctttaactgc tgtcctcatt aggtaatgat aaatatttcc cttaaataat tgactatttt 4800
gctgtgtttt aaaaatgatt gaaatttatc ttgccatata ttaataaggag cacatttagt tgaggtagaa 4920
tgactgagct aatccttgaga atatatctgt aaaataggag tttagctttag tgaatttcaa aagtaattggg 4980
aaggtaggac tcttagacaaa accttctatt ttagttagctt gaaagagctt aatcatatgc agtaagtatt 5040
tcttgaggta tagattttta ttagtagctt gaaaagagctt tttatcctag ggccaagtgt 5100
tttattacca ataaatttta aattttttaa gaaaaatatt tttatcctag accctaacag ttttaccacc 5160
tgcctgccac caatcagtaa gtttagtctat aacaaatttt atagaaaagtc actccttggc aaaagtgtta 5220
tagtaacagt catttctgaa aatatgttgg atagaaaagtc atggcatcta tcttgaaagt aatccttgat 5280
gaatttgctt ttgtgccatc tattcctttt aatttagaaa ttaacatgat atcttaaat acccttatga 5340
tggagattga aagatgctgt aatttagaaa catagatttt ccttcaaaaa atgaacattt atatatctac 5400
aatatagttt tgtataatag tttgaaagcc tactttctga agaaaatggg gggatttttt 5460
aaaaatatgg tttatcatga ttaaatatca aaaaattgcc ctatgaaaac tttaaatctc taaaacattt 5520
gaaatactac catattttgtg attttattgag aataaaaaatc cattttgaaa tgtaaaattt 5580
ttatgatctg attcagtttt aagaaaacat gaatgaacta gaagatatta aaaacatttg 5640
acattggtaa gaaatattga tactgatatt gatttttata taggtattta tttcagaatt 5700
gataatttga gaaaaataca tgtgagtcac tttttctgtt tctcttttct cttaacgatt 5760
atcactgtaa ttctgaatct gaaaggtaaa acaattagtc aaaatattat tgccatcatt 5820
ctacctgtgt tatgaaacta cttattcata gtttaattctc attaacactt acatttccat 5880
aaagaaaact caagtattaa taaaagagac tttactggct taagagggct gtgaaagatt 5940
tttgatagtg aatcatgacc ctaagggaga gatttggtgt ataaaagtat tgtatataat 6000
agatcagcga tttttgtaag gcaaacagaa tttgtaagtt ggcagatctt cctaagttgc 6060
aaaatgtaat gatgaccttg gtggagaaag atgagtcgtt cttggaatac ctatgtgcag 6120
ccactacca tctcaatgtc acctgttttg cattcttgga tagcttgat atgtagtagt 6180
ttgatgaata atttaaagaa aaacacctaa aatttgaaaa atgattgtag gatcaaaaaa 6240
ggcagatgaa attacttaat actcagtggt ttggagagta ttccttttag tttgttggtt 6300
ggctggtttg aacgatagaa atatgcagca tgcaatatat gcttatattt cattttaatt 6360
tctgatatat aatgaacttc ttgggagagg tactgaatct ttgatgttt ttgtcattgt 6420
tctcaagtgc aatataacaa tgtaaccaaa tctagataat ttcaaagttg tcattaattt 6480
agtaagccta atataaacia atatttgtat tatttttgtt agcaggaaaag agtgattaag 6540
tgaggttatt tacccttaaa tgggtccattc tgcattgtat ttcaggctgg aaatgaatta 6600
ttctttacca gttttgaaac actttgaaat atcctaaggt aacttggaag ctgtgtagta 6660
tatcaaatta atttgctacc taataacata gaaagtaaat atctttgtgg tcaccacat 6720
tgggtgagac agaaaatgaa tctgttctaa aatttgaat ttgctaactt gatttgagtt 6780
agtgaaaact ggtacagtgt tctgcttgat ttacaacatg taacttgtga ctgtacaata 6840
aacataagca tatggtaaaa aaaaaaaaaa a

```

<210> 74

<211> 6696

<212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 7505011CB1

<400> 74
 ttcttttcaag aagatcaggg acaactgatt tgaagtctac tctgtgcttc taaatcccca 60
 attctgtctga aagttagata ccctagagcc ctagagcccc agcagcacc agccaaacc 120
 acctccacca tggggggccat gactcagctg ttggcaggtg tctttcttgc tttccttgcc 180
 ctcgctaccg aaggtgggggt cctcaagaaa gtcattccggc acaagcgaca gagtgggggtg 240
 aacgccaccc tgccagaaga gaaccagcca gtgggtgttta accacgttta caacatcaag 300
 ctgccagtgg gatcccagtg ttccgggtgat ctggagtcag ccagtgggga gaaagacctg 360
 gcaccgcctt cacagcccag cgaaagcttt caggagcaca cagtggatgg ggaaaaccag 420
 attgtcttca cactcgcac caacatcccc cgccgggcct gtggctgtgc cgcagccccct 480
 gatgttaagg agctgctgag cagactggag gagctggaga acctgggtgtc ttccctgagg 540
 gagcaatgta ctgcaggagc aggtgtgtgt ctccagcctg ccacaggccg cttggacacc 600
 aggccttct gtacgggtcg gggcaacttc agcactgaag gatgtggctg tgtctgcgaa 660
 cctggctgga aaggcccaa ctgctctgag cccgaatgtc caggcaactg tcaccttca 720
 ggccggtgca ttgatgggca gtgcatctgt gacgacggct tcacggcgca ggcagtcagc 780
 cagctggctt gcccagcga ctgcaatgac caggcaagt gcgtgaatgg agtctgcac 840
 tgtttcgaag gctacgccgg ggctgactgc agccgtgaaa tctgcccagt gccctgcagt 900
 gaggagcacg gcacatgtgt agatggcctt tgtgtgtgcc acgatggctt tgcaggcgat 960
 gactgcaaca agcctctgtg tctcaacaat tgctacaacc gtggacgatg cgtggagaat 1020
 gagtgcgtgt gtgatgaggg ttccacgggc gaagactgca gtgagctcat ctgccccaat 1080
 gactgcttcg accggggccg ctgcatcaat ggcacctgtc actgcgaaga aggcctcaca 1140
 ggtgaagact cggggaaacc cacctgcccc catgctgtcc acaccagggt ccggtgtgag 1200
 gaggggcagt gtgtatgtga tgagggtttt gccggtgtgg actgcagcga gaagaggtgt 1260
 cctgctgact gtcacaatcg tggccgctgt gtagacgggc ggtgtgagtg tgatgatggt 1320
 ttcactggag ctgactgtgg ggagctcaag tgtcccaatg gctgcagtgg ccatggccgc 1380
 tgtgtcaatg ggcagtgtgt gtgtgatgag ggctatactg gggaggactg cagccagcta 1440
 cgggtgcccc atgactgtca cagtcggggc cgctgtgtcg agggcaaatg tgtatgtgag 1500
 caaggtctca agggctatga ctgcagtgaac atgagctgcc ctaatgactg tcaccagcac 1560
 ggccgctgtg tgaatggcat gtgtgtttgt gatgacggct acacagggga agactgccgg 1620
 gatcgccaat gccccaggga ctgcagcaac aggggcctct gtgtggacgg acagtgcgtc 1680
 tgtgaggacg gcttcaccgg cctgactgtg gcagaactct cctgtccaaa tgactgccat 1740
 ggccgggggtc gctgtgtgaa tgggcagtgc gtgtgccatg aaggatttat gggcaaagac 1800
 tgcaaggagc aaagatgtcc cagtgactgt catggccagg gccgtgtcgt ggcaggccag 1860
 tgcattctgc acgagggtt cagaggcctg cactgtggcc agcactcctg cccagtgac 1920
 tgcaacaact taggacaatg cgtctcgggc cgctgcatct gcaacgaggg ctacagcgga 1980
 gaagactgct cagaggtgtc tctcccaaaa gacctcgttg tgacagaagt gacggaagag 2040
 acggtcaacc tggcctggga caatgagatg cgggtcacag agtaccttgt cgtgtacacg 2100
 cccaccacag aggggtgtct ggaaatgcag ttccgtgtgc ctggggacca gacgtccacc 2160
 atcatccagg agctggaacc tgggtgtggag tactttatcc gtgtatttgc catcctggag 2220
 aacaagaaga gcattcctgt cagcgccag gtggccagct acttacctgc acctgaaggc 2280
 ctgaaattca agtccatcaa ggagacatct gtggaagtgg agtgggatcc tctagacatt 2340
 gcttttga aa cctgggagat catcttccgg aatatgaata aagaagatga gggagagatc 2400
 accaaaagcc tgaggaggcc agagacctct taccggcaaa ctggtctagc tcctgggcaa 2460
 gagtatgaga tatctctgca catagtga aa aacaataccc ggggccttg cctgaagagg 2520
 gtgaccacca cagccttgga tgccccagc cagatcgagg tgaaagatgt cacagacacc 2580
 actgccttga tcacctggt caagcccctg gctgagatcg atggcattga gctgacctac 2640
 ggcatcaaa acgtgccagg agaccgtacc accatcgatc tcacagagga cgagaaccag 2700
 tactccatcg ggaacctgaa gcctgacact gagtacgagg tgtccctcat ctcccgcaga 2760
 ggtgacatgt caagcaacc agccaaagag accttcacaa caggcctcga tgctcccagg 2820
 aatcttcgac gtgtttccca gacagataac agcatcacc tggaatggag gaatggcaag 2880
 gcagctattg acagttacag aattaagtat gccccatct ctggagggga ccacgtgag 2940
 gttgatgttc caaagagcca acaagccaa accaaaacca cactcacagg cctgaggccg 3000
 ggaactgaat agtttctgtc gtgaaggaa acaaggagag caatccagcg 3060
 accatcaacg cagccacaga gttggacacg cccaaggacc ttcagggttc tgaaactgca 3120
 gagaccagcc tgacctgtct ctggaagaca ccgttggcca aatttgaccg ctaccgcctc 3180
 aattacagtc tccccacagg ccagtgggtg ggagtgcagc ttccaagaaa caccacttcc 3240
 tatgtcctga gaggcctgga accaggacag gagtacaatg tctcctgac agccgagaaa 3300

| | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ggcagacaca | agagcaagcc | cgcacgtgtg | aaggcatcca | ctgaacgagc | ccctgagctg | 3360 |
| gaaaacctca | ccgtgactga | ggttggctgg | gatggcctca | gactcaactg | gaccgcagct | 3420 |
| gaccaggcct | atgagcactt | tatcattcag | gtgcaggagg | ccaacaaggt | ggaggcagct | 3480 |
| cggaacctca | ccgtgcctgg | cagccttcgg | gctgtggaca | taccgggcct | caaggctgct | 3540 |
| acgccttata | cagtctccat | ctatgggtcg | ttccagggtc | atagaacacc | agtgtctctt | 3600 |
| gctgaggcct | ccacagggga | aactcccaat | ttgggagagg | tcgtgggtggc | cgagggtgggc | 3660 |
| tgggatgccc | tcaaaactca | ctggactgct | ccagaagggg | cctatgagta | ctttttcatt | 3720 |
| caggtgcagg | aggctgacac | agtagaggca | gcccagaacc | tcaccgtccc | aggaggactg | 3780 |
| aggtccacag | acctgcctgg | gctcaaaagca | gccactcatt | ataccatcac | catccgcggg | 3840 |
| gtcactcagg | acttcagcac | aacccctctc | tctgttgaag | tcttgacaga | ggagggttcca | 3900 |
| gatatgggaa | acctcacagt | gaccgagggt | agctgggatg | ctctcagact | gaactggacc | 3960 |
| acgccagatg | gaacctatga | ccagtttact | attcagggtcc | aggaggctga | ccagggtggaa | 4020 |
| gaggctcaca | atctcacggg | tcctggcagc | ctgcgttcca | tggaaatccc | aggcctcagg | 4080 |
| gctggcagtc | cttacacagt | caccctgcac | ggcgagggtca | ggggccacag | cactgcagcc | 4140 |
| cttgctgtag | aggtcgctac | agaggatctc | ccacagctgg | gagatttagc | cgtgtctgag | 4200 |
| gttggctggg | atggcctcag | actcaactgg | accgcagctg | acaatgccta | tgagcacttt | 4260 |
| gtcattcagg | tgcaggagggt | caacaaagtg | gaggcagccc | agaacctcac | gttgcctggc | 4320 |
| agcctcaggg | ctgtggacat | cccggggcctc | gaggctgcca | cgcttatatg | agtctccatc | 4380 |
| tatgggggtga | tccgggggcta | tagaacacca | gtactctctg | ctgaggcctc | cacagccaaa | 4440 |
| gaacctgaaa | ttggaaactt | aaatgttttc | gacataactc | ccgagagctt | caatctctcc | 4500 |
| tggatggcta | ccgatgggat | cttcgagacc | tttaccattg | aaattattga | ttccaatagg | 4560 |
| ttgctggaga | ctgtggaata | taatatctct | ggtgctgaac | gaactgcccc | tatctcaggg | 4620 |
| ctaccccccta | gtactgattt | tattgtctac | ctctctggac | ttgctcccag | catccggacc | 4680 |
| aaaacctatca | gtgccacagc | cacgacagaa | gccgaaccgg | aagttgacaa | ccttctgggt | 4740 |
| tcagatgcca | ccccagacgg | tttccgtctg | tcctggacag | ctgatgaagg | ggtcttcgac | 4800 |
| aattttgttc | tcaaaactcag | agataccaaa | aagcagctcg | agccactgga | aataacccta | 4860 |
| cttgcctcccg | aacgtaccag | ggacttaaca | ggtctcagag | aggctactga | atacgaaatt | 4920 |
| gaactctatg | gaataagcaa | aggaaggcga | tcccagacag | tcagtgtctat | agcaacaaca | 4980 |
| gccatgggct | ccccaaagga | agtcattttc | tcagacatca | ctgaaaattc | ggctactgtc | 5040 |
| agctggaggg | caccacacagc | ccaagtggag | agcttccgga | ttacctatgt | gccattaca | 5100 |
| ggaggtagac | cctccatggg | aactgtggac | ggaaccaaga | ctcagaccag | gctgggtgaa | 5160 |
| ctcatacctg | gcgtggagta | ccttgtcagc | atcatcgcca | tgaagggtct | tgaggaaagt | 5220 |
| gaacctgtct | cagggtcatt | caccacagct | ctggatggcc | catctggcct | ggtgacagcc | 5280 |
| aacatcactg | actcagaagc | cttggccagg | tggcagccag | ccattgccac | tgtggcagct | 5340 |
| tatgtcatct | cctacacagg | cgagaaagtg | ccagaaatta | cacgcacggg | gtccgggaac | 5400 |
| acagtggagt | atgctctgac | cgacctcgag | cctgccacgg | aatacacact | gagaatcttt | 5460 |
| gcagagaaaag | ggccccagaa | gagctcaacc | atcactgcca | agttcacac | agacctcgat | 5520 |
| tctccaagag | acttgactgc | tactgagggt | cagtcggaaa | ctgcccctct | tacctggcga | 5580 |
| cccccccg | catcagtcac | cggttacctg | ctggtctatg | aatcagtgga | tggcacagtc | 5640 |
| aagggaagtca | ttgtgggtcc | agataccacc | tcctacagcc | tggcagacct | gagcccatcc | 5700 |
| accactaca | cagccaagat | ccaggcactc | aatggggccc | tgaggagcaa | tatgatccag | 5760 |
| accatcttca | ccacaatttg | actcctgtac | cccttcccc | aggactgctc | ccaagcaatg | 5820 |
| ctgaatggag | acacgacctc | tggcctctac | accattttatc | tgaatggtga | taagggtcag | 5880 |
| gcgctggaag | tcttctgtga | catgacctct | gatgggggtg | gatggattgt | gttcctgaga | 5940 |
| cgcaaaaacg | gacgcgagaa | cttctaccaa | aactggaagg | catatgctgc | tggatttggg | 6000 |
| gaccgcagag | aagaattctg | gcttgggctg | gacaacctga | acaaaatcac | agcccagggg | 6060 |
| cagtacgagc | tccgggtgga | cctgcggggac | catggggaga | cagcctttgc | tgtctatgac | 6120 |
| aagttcagcg | tgggagatgc | caagactcgc | tacaagctga | aggtggaggg | gtacagtggg | 6180 |
| acagcagggtg | actccatggc | ctaccacaat | ggcagatcct | tctccacctt | tgacaaggac | 6240 |
| acagattcag | ccatcaccaa | ctgtgctctg | tcctacaaag | gggctttctg | gtacaggaac | 6300 |
| tgtcacccgtg | tcaacctgat | ggggagatat | ggggacaata | accacagtca | gggcgttaac | 6360 |
| tggttccact | ggaagggcca | cgaacactca | atccagtttg | ctgagatgaa | gctgagacca | 6420 |
| agcaacttca | gaaatcttga | aggcaggcgc | aaacgggcat | aaattccagg | gaccactggg | 6480 |
| tgagagagga | ataaggccca | gagcgaggaa | aggattttac | caaagcatca | atacaaccag | 6540 |
| cccaaccatc | ggtccacacc | tgggcatttg | gtgagagtca | aagctgacca | tggatccctg | 6600 |
| gggccaaacgg | caacagcatg | ggcctcacct | cctctgtgat | ttctttcttt | gcaccaaaga | 6660 |
| catcagtctc | caacatgttt | ctgttttgtt | gtttga | | | 6696 |

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 June 2003 (12.06.2003)

PCT

(10) International Publication Number
WO 03/047526 A3

(51) International Patent Classification⁷: C12N 1/21, 5/12, 1/20, 15/63, C07H 21/04, C07K 14/00

(21) International Application Number: PCT/US02/38437

(22) International Filing Date:
26 November 2002 (26.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

| | | |
|------------|-------------------------------|----|
| 60/334,343 | 30 November 2001 (30.11.2001) | US |
| 60/340,278 | 7 December 2001 (07.12.2001) | US |
| 60/345,069 | 4 January 2002 (04.01.2002) | US |
| 60/351,352 | 25 January 2002 (25.01.2002) | US |
| 60/357,168 | 14 February 2002 (14.02.2002) | US |
| 60/369,128 | 29 March 2002 (29.03.2002) | US |
| 60/370,802 | 5 April 2002 (05.04.2002) | US |

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). BECHA, Shanya, D. [US/US]; 21062 Gary Drive # 117, Castro Valley, CA 94546 (US). BHATIA, Umesh [US/US]; 5212 Union Avenue, San Jose, CA 95124 (US). BLAKE, Julie, J. [US/US]; 3818 Pacheco Street, San Francisco, CA 94116 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). BURRILL, John, D. [US/US]; 2218 Brewster Avenue, Redwood City, CA 94062 (US). DELEGEANE, Angelo, M. [US/US]; 594 Angus Drive, Milpitas, CA 95035 (US). ELLIOTT, Vicki, S. [US/US]; 3770 Polton Place Way, San Jose, CA 95121 (US). GANDHI, Ameena, R. [US/US]; 705 5th Avenue, San Francisco, CA 94118 (US). GIETZEN, Kimberly, J. [US/US]; 691 Los Huecos Drive, San Jose, CA 95123 (US). GORVAD, Ann, E. [US/US]; 369 Marie Common, Livermore, CA 94550 (US). GRIFFIN, Jennifer, A. [US/US]; 33691 Mello Way, Fremont, CA 94555 (US). HO, Anne [FR/US]; 1279 Poplar Avenue, #114, Sunnyvale, CA 94086 (US). JIN, Pei [US/US]; 320 Curtner Avenue #D, Palo Alto, CA 94306 (US). KABLE, Amy, E. [US/US]; 2345 Polk Street #4, San Francisco, CA 94109 (US). LAL, Preeti, G. [US/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). LEE, Ernestine, A. [US/US]; 20523 Crow Creek Road,

Castro Valley, CA 94552 (US). LEE, Sally [US/US]; 3643 26th Street, San Francisco, CA 94110 (US). LEE, Soo, Yeun [KR/US]; 40 Westdale Avenue, Daly City, CA 94015 (US). MARQUIS, Joseph, P. [US/US]; 4428 Lazy Lane, San Jose, CA 95135 (US). LEHR-MASON, Patricia, M. [US/US]; 360 Clarke Lane, Morgan Hill, CA 95014 (US). RAMKUMAR, Jayalaxmi [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). RICHARDSON, Thomas, W. [US/US]; 616 Canyon Road #107, Redwood City, CA 94062 (US). SPRAGUE, William, W. [US/US]; 611 13th Street # C, Sacramento, CA 95814 (US). SWARNAKAR, Anita [CA/US]; 8 Locksley Avenue #5D, San Francisco, CA 94122 (US). TANG, Tom, Y. [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). TRAN, Bao [US/US]; 750 Salberg Avenue, Santa Clara, CA 95051 (US). TRAN, Uyen, K. [US/US]; 2638 Mabury Square, San Jose, CA 95133 (US). CHAWLA, Narinder, K. [US/US]; 33 Union Square, #712, Union City, CA 94587 (US). WARREN, Bridget, A. [US/US]; 1810 S. El Camino Real #B103, Encinitas, CA 94024 (US). XU, Yuming [US/US]; 1739 Walnut Drive, Mountain View, CA 94040 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). ZHENG, Wenjin [CN/US]; 9 Sutter Creek Lane, Mountain View, CA 94043 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report

(88) Date of publication of the international search report:
13 November 2003

[Continued on next page]

(54) Title: CELL ADHESION AND EXTRACELLULAR MATRIX PROTEINS

(57) Abstract: Various embodiments of the invention provide human cell adhesion and extracellular matrix proteins (CADECM) and polynucleotides which identify and encode CADECM. Embodiments of the invention also provide expression vectors, host cells, antibodies, agonists, and antagonists. Other embodiments provide methods for diagnosing, treating, or preventing disorders associated with aberrant expression of CADECM.



WO 03/047526 A3



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/38437

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 1/21, 5/12, 1/20, 15/63; C07H 21/04 C07K 14/00
US CL : 435/69.1, 325, 252.3, 320.1; 536/23.5; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 325, 252.3, 320.1; 536/23.5; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
STN, WEST, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|---|
| X | LEUNG, E. et al. Cloning of the mucosal addressin MAdCAM-1 from human brain: identification of novel alternatively spliced transcripts. Immunol Cell Biol. December 1996, Vol. 74, No. 6, pages 490-496, see entire document. | 1, 3, 6-7, 9, 12-15 and 17 |
| X | SHYJAN et al. Human mucosal addressin cell adhesion molecule-1 (MAdCAM-1) demonstrates structural and functional similarities to the alpha 4 beta 7-integrin binding domains of murine MAdCAM-1, but extreme divergence of mucin-like sequences. J. Immunol. 15 April 1996 (15.4.1996), Vol. 156, No. 8, pages 2851-2857, see entire document. | 1, 3, 6-7, 9, 12-15 and 17 |
| X | US 6,037,324 A (SCHWEDER et al). 14 March 2000 (14.3.2000), see SEQ ID NO: 67-68. | 1, 3, 6-7, 9, 12-15 and 17 |
| A | WO/9624673 A1 (LEUKOSITE INC.) 01 August 1996 (15.08.1996). see claim 5, page 104-106. | 2, 4-5, 56 and 93 |
| X | US 5,962, 643 A (SHEPPARD et al) 05 October 1999 (05.10.1999). SEQ ID NO: 26-27 | 1, 3, 6-7, 9, 12-15 and 17 |
| A | Database GenBank, Accession Number AL831998. 13 May 2003 (13.5.2003). | 1-7, 9-10, 12-15, 17-18, 22, 77 and 114 |

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

19 May 2003 (19.05.2003)

Date of mailing of the international search report

17 JUL 2003

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450

Facsimile No. 703 305-3230

Authorized officer

Maher M. Haddad

Telephone No. 703 308-0196

INTERNATIONAL SEARCH REPORT

PCT/US02/38437

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|----------------------------|
| X | SHEPPARD et al. Complete amino acid sequence of a novel integrin beta subunit (beta 6) identified in epithelial cells using the polymerase chain reaction. J Biol Chem. 15 July 1990 (15.7.1990), Vol. 265, No. 20, page 11502-11507, see entire document. | 1, 3, 6-7, 9, 12-15 and 17 |

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/38437

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1-7, 9-10, 12-15, 1-18, 56, 77, 93 and 114
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US02/38437

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This International Search Authority has found 777 inventions claimed in the International Application covered by the claims indicated below:

1. Claims 1-7, 9-10, 12-15, 17-18, 56 and 93, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 1, the nucleic acid of SEQ ID NO: 38, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
2. Claims 1-7, 9-10, 12-15, 17-18, 57 and 94, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 2, the nucleic acid of SEQ ID NO: 39, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
3. Claims 1-7, 9-10, 12-15, 17-18, 58 and 95, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 3, the nucleic acid of SEQ ID NO: 40, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
4. Claims 1-7, 9-10, 12-15, 17-18, 59 and 96, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 4, the nucleic acid of SEQ ID NO: 41, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
5. Claims 1-7, 9-10, 12-15, 17-18, 60 and 97, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 5, the nucleic acid of SEQ ID NO: 42, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
6. Claims 1-7, 9-10, 12-15, 17-18, 61 and 98, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 6, the nucleic acid of SEQ ID NO: 43, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
7. Claims 1-7, 9-10, 12-15, 17-18, 62 and 99, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 7, the nucleic acid of SEQ ID NO: 44, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
8. Claims 1-7, 9-10, 12-15, 17-18, 63 and 100, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 8, the nucleic acid of SEQ ID NO: 45, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
9. Claims 1-7, 9-10, 12-15, 17-18, 64 and 101, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 9, the nucleic acid of SEQ ID NO: 46, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
10. Claims 1-7, 9-10, 12-15, 17-18, 65 and 102, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 10, the nucleic acid of SEQ ID NO: 47, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
11. Claims 1-7, 9-10, 12-15, 17-18, 66 and 103, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 11, the nucleic acid of SEQ ID NO: 48, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.
12. Claims 1-7, 9-10, 12-15, 17-18, 67 and 104, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 12, the nucleic acid of SEQ ID NO: 49, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

PCT/US02/38437

- Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

PCT/US02/38437

28. Claims 1-7, 9-10, 12-15, 17-18, 83 and 120, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 28, the nucleic acid of SEQ ID NO: 65, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

29. Claims 1-7, 9-10, 12-15, 17-18, 84 and 121, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 29, the nucleic acid of SEQ ID NO: 66, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

30. Claims 1-7, 9-10, 12-15, 17-18, 85 and 122, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 30, the nucleic acid of SEQ ID NO: 67, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

31. Claims 1-7, 9-10, 12-15, 17-18, and 86 and 123, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 31, the nucleic acid of SEQ ID NO: 68, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

32. Claims 1-7, 9-10, 12-15, 17-18, 87 and 124, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 32, the nucleic acid of SEQ ID NO: 69, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

33. Claims 1-7, 9-10, 12-15, 17-18, 88 and 125, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 33, the nucleic acid of SEQ ID NO: 70, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

34. Claims 1-7, 9-10, 12-15, 17-18, 89 and 126, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 34, the nucleic acid of SEQ ID NO: 71, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

35. Claims 1-7, 9-10, 12-15, 17-18, 90 and 127, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 35, the nucleic acid of SEQ ID NO: 72, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

36. Claims 1-7, 9-10, 12-15, 17-18, 91 and 128, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 36, the nucleic acid of SEQ ID NO: 73, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

37. Claims 1-7, 9-10, 12-15, 17-18, 92 and 129, drawn to a polypeptide comprising an amino acid sequence of SEQ ID NO: 37, the nucleic acid of SEQ ID NO: 74, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

38-74. Claim 8, drawn to a transgenic organism comprising a recombinant polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY.

75-111. Claims 11, 31-32, 34, 36-43, drawn to an isolated antibody which specifically binds to a polypeptide of SEQ ID NO: 1-36 RESPECTIVELY.

112-148. Claim 16, drawn to a method of detecting a target polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY, in a sample comprising amplifying said target polynucleotide or fragment thereof using polymerase chain reaction.

149-185. Claim 19, drawn to a method of treating a disease or condition associated with the decreased expression of functional CADECM, comprising administering to a patient a composition comprising a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

186-222. Claim 20, drawn to a method of screening a compound for effectiveness as an agonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

223-259. Claim 21, drawn to a composition comprising an agonist compound identified by a method of screening a compound for effectiveness as an agonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

260-296. Claim 22, drawn to a method for treating a disease or condition associated with decreased expression of functional CADECM, comprising administering a composition comprising an agonist compound identified by a method of screening a compound for effectiveness as an agonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

INTERNATIONAL SEARCH REPORT

PCT/US02/38437

297-333. Claim 23, drawn to a method of screening a compound for effectiveness as an antagonist of a polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

334-370. Claim 24, drawn to a composition comprising an antagonist compound identified by a method of screening a compound for effectiveness as an antagonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

371-407. Claim 25, drawn to a method for treating a disease or condition associated with decreased expression of functional CADECM, comprising administering a composition comprising an antagonist compound identified by a method of screening a compound for effectiveness as an antagonist of a polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

408-444. Claim 26, drawn to a method of screening for a compound that specifically binds to the polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

445-481. Claim 27, drawn to a method of screening for a compound that modulates the activity of the polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

482-518. Claim 28, drawn to a method of screening for a compound for effectiveness in altering expression of a target polynucleotide of SEQ ID NO:38-74 RESPECTIVELY.

519-555. Claim 29, drawn to a method of assessing toxicity of a test compound comprising hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY.

556-592. Claim 30, drawn to a method for a diagnostic test for a condition or disease associated with the expression of CADECM in a sample, comprising combining the biological sample with an antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

593-629. Claims 33 and 35, drawn to a method of diagnosing a condition or disease associated with the expression of CADECM in a subject, comprising administering a composition comprising an antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

630-666. Claim 44, drawn to a method of detecting a polypeptide comprising an amino of SEQ ID NO: 1-37 RESPECTIVELY, comprising incubating the antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY with the sample.

667-703. Claim 45, drawn to a method of purifying a polypeptide comprising an amino acid sequence of SEQ ID NO:1-37 RESPECTIVELY, using the antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

704-740. Claims 46, 48-55, drawn to a microarray wherein at least one element of the microarray is a polynucleotide of SEQ ID NO:38-74 RESPECTIVELY.

741-777. Claim 47, drawn to a method of generating an expression profile of a sample which contains polynucleotides comprising contacting the elements of the microarray wherein at least one element of the microarray is a polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY, with the labeled polynucleotides of the sample.

The inventions listed as Group 1-777 do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of the Group.

The specific technical feature of the Group 1 invention is the polypeptide comprising an amino acid sequence of SEQ ID NO: 1, the nucleic acid of SEQ ID NO: 38, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe claimed therein while the specific technical feature of the Group 2 invention is the polypeptide comprising an amino acid sequence of SEQ ID NO: 2, the nucleic acid of SEQ ID NO: 39, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe claimed therein.

The specific technical feature of the Groups 3-37 inventions are the polypeptide comprising an amino acid sequence of SEQ ID NO: 2-37 Respectively, the nucleic acid of SEQ ID NO: 39-74 Respectively, a recombinant polynucleotide, host cells, methods of producing the polypeptide and a method of detecting a target polynucleotide in a sample comprising hybridizing the sample with a probe.

The specific technical feature of the Groups 38-74 inventions are the transgenic organism comprising a recombinant polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY.

INTERNATIONAL SEARCH REPORT

PCT/US02/38437

The specific technical feature of the Groups 75-111 inventions are the isolated antibody which specifically binds to a polypeptide of SEQ ID NO: 1-36 RESPECTIVELY.

The specific technical feature of the Groups 112-148 inventions are the method of detecting a target polynucleotide of SEQ ID NO: 38-74, RESPECTIVELY, in a sample comprising amplifying said target polynucleotide or fragment thereof using polymerase chain reaction.

The specific technical feature of the Groups 149-185 inventions are the method of treating a disease or condition associated with the decreased expression of functional CADECM, comprising administering to a patient a composition comprising a polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

The specific technical feature of the Groups 186-222 inventions are the method of screening a compound for effectiveness as an agonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 223-259 inventions are the composition comprising an agonist compound identified by a method of screening a compound for effectiveness as an agonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 260-296 inventions are the method for treating a disease or condition associated with decreased expression of functional CADECM, comprising administering a composition comprising an agonist compound identified by a method of screening a compound for effectiveness as an agonist of a polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

The specific technical feature of the Groups 297-333 inventions are the method of screening a compound for effectiveness as an antagonist of a polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

The specific technical feature of the Groups 334-370 invention is the composition comprising an antagonist compound identified by a method of screening a compound for effectiveness as an antagonist of a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 371-407 inventions are the method for treating a disease or condition associated with decreased expression of functional CADECM, comprising administering a composition comprising an antagonist compound identified by a method of screening a compound for effectiveness as an antagonist of a polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

The specific technical feature of the Groups 408-444 inventions are the method of screening for a compound that specifically binds to the polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 445-481 inventions are the method of screening for a compound that modulates the activity of the polypeptide of SEQ ID NO:1-37 RESPECTIVELY.

The specific technical feature of the Groups 482-518 inventions are the method of screening for a compound for effectiveness in altering expression of a target polynucleotide of SEQ ID NO:38-74 RESPECTIVELY.

The specific technical feature of the Groups 519-555 inventions are the method of assessing toxicity of a test compound comprising hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY.

The specific technical feature of the Groups 556-592 inventions are the method for a diagnostic test for a condition or disease associated with the expression of CADECM in a sample, comprising combining the biological sample with an antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 593-629 inventions are the method of diagnosing a condition or disease associated with the expression of CADECM in a subject, comprising administering a composition comprising an antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 630-666 inventions are the method of detecting a polypeptide comprising an amino of SEQ ID NO: 1-37 RESPECTIVELY, comprising incubating the antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY with the sample.

The specific technical feature of the Groups 667-703 inventions are the method of purifying a polypeptide comprising an amino acid sequence of SEQ ID NO:1-37 RESPECTIVELY, using the antibody which specifically binds to a polypeptide of SEQ ID NO: 1-37 RESPECTIVELY.

The specific technical feature of the Groups 704-740 inventions are the microarray wherein at least one element of the microarray is a polynucleotide of SEQ ID NO:38-74 RESPECTIVELY.

INTERNATIONAL SEARCH REPORT

PCT/US02/38437

The specific technical feature of the Groups 741-777 inventions are the method of generating an expression profile of a sample which contains polynucleotides comprising contacting the elements of the microarray wherein at least one element of the microarray is a polynucleotide of SEQ ID NO: 38-74 RESPECTIVELY, with the labeled polynucleotides of the sample.

Since the special technical feature of the Group 1 invention is not present in the Groups 2-777 claims and the special technical feature of the Groups 2-777 inventions are not present in the Group 1 claims, unity of invention is lacking